New York City Payroll - Project Proposal - Group 6

Examining trends in payroll funds allocation for employees of New York City

Nicole Cosmany
University of Colorado
Boulder
Boulder, CO USA
nico3601@colorado.edu

Ebrahim Azarisooreh
University of Colorado
Boulder
Boulder, CO USA
ebaz7868@colorado.edu

Problem State/Motivation

Taxpayers in the United States fund a wide range of government services ranging from payroll to defense to infrastructure and more. Naturally, they want to know how and where their money is being spent. To ensure transparency and accountability, government organizations make this data available online for public viewing and use.

In this project, we are focusing specifically on New York City's payroll data. Our goal is to explore how the city allocates its payroll budget across various job titles and departments. By analyzing this data, we hope to gain a deeper understanding of budget allocation trends, identify patterns over time, and uncover insights into the relative financial priorities of various departments and roles within the city.

We propose the following questions:

1. **Allocation:** How is the total NYC payroll budget allocated among departments? Among boroughs?
2. **Pay**: How does pay for a certain job title vary from year to year? Does the average pay for a certain job title vary noticeably across boroughs?
3. **Overtime**: How much of the total payroll spending is for overtime? Are certain departments consistently spending more on overtime? Is there a relationship between the number of employees in a department and overtime spending? Are certain job titles working overtime more often than others?
4. **Duration of Employment:**

   How does pay rate/salary relate to employment duration? Do employees who are paid more/less have longer/shorter tenures than others? How does this vary across boroughs, departments, and job title?

Ultimately, this analysis will reveal key insights into how government resources are prioritized, identify how pay rate is affected by time and place, and the impact of overtime spending on the overall budget. Our results could be relevant to a wide range of stakeholders including New York City taxpayers concerned with allocation, budget committees planning for future years, hiring managers looking to address overtime concerns, or even city employees looking to understand what factors contribute to their pay rates.

Literature Survey

Overtime

New York City has been seeing consistent increases in overtime spending over the last decade but no adjustments have been made to the budget to account for this issue, thus the city is consistently surpassing its proposed budget year and after (1). This is problematic because it defeats the purpose of having a budget if it is constantly overspent. Police are generally the

biggest overspenders. They had a period from 2015 - 2019 when overtime spending in the police department stabilized, but it has since continued to grow again(1), according to the article published in 2023. With our dataset ranging from 2014-2024 we expect to be able to confirm these previously observed trends. In an attempt to mediate overspending, overtime budgets were reduced but this has been shown to be ineffective as budgets continue to be surpassed(1).

(1)https://comptroller.nyc.gov/reports/overtime-overview/#:~:text=The%20FY%202022%20actual%20overtime,uniformed%20overtime%20in%20FY%202022

Wage Trends

According to Forbes, employees tend to be rewarded more often for changing jobs rather than remaining loyal. On the other hand, many employees have the nagging concern that changing jobs too often can reflect negatively on their resumes and perception of work history. We'd like to know if there's a relationship between base pay rate and employment duration.

(2)https://www.forbes.com/sites/cameronkeng/2014/06/22/employees-that-stay-in-companies-longer-than-2-years-get-paid-50-less/

Proposed Work

We plan to address the questions raised in our Motivation Statement. To answer each question, we will perform the following calculations:

1. Allocation
    a. calculate total budget for each year by adding total salary
    b. sum total compensation, group by borough
    c. sum total compensation, group by department
2. Pay
    a. sort unique job titles and pick titles of interest (Police Officer, Firefighter, etc.)
    b. calculate average salary per job title
    c. calculate average salary per job title and group by borough
    d. line chart showing average salary vs. time for chosen job titles
    e. bar chart showing average salary per borough for chosen job titles
3. Overtime
    a. Determine average number of OT hours worked by job over the years
    b. Graph OT hours worked by job title.
    c. Determine if a relationship exists between OT hours worked over the years in the dataset.
4. Duration of Employment
    a. Calculate average employment duration by job
    b. Use calculated salary average for job titles and graph them by employment duration
    c. Graph employment duration by job-title and salary and determine if a correlation exists on any of these item-pairs.

Missing values: If missing salary values come up, we will impute by using an average of other employees in the same department with the same title in the same year. Most other missing values can be ignored - if we don't know the job title or department, there's not much we can do about that besides remove it from our calculations. We do anticipate many, if any, missing values.

Compensation Calculations: Not all pay information is presented as an annual salary. We plan to analyze all pay as annual salaries. This will require calculating salary for hourly or per diem employees using pay basis, regular pay,

regular hours, and additional compensation attributes.

Reduce Redundancy: There is currently some redundancy among department names, where certain departments are listed with sub departments or slight differences in names. We aim to rectify these by standardizing names across departments. We also plan to drop the Payroll Number column since this is another way of identifying departments.

Data Set:

Available online at: https://data.cityofnewyork.us/City-Government/Citywide-Payroll-Data-Fiscal-Year-/k397-673e/about_data

Additionally, both group members have the dataset downloaded.

Some data is omitted in certain departments due to confidentiality policies.

The dataset has 6.22 million entries, where each entry corresponds to a single employee's pay information for the given fiscal year. Each entry has 17 attributes, which we define below:

Fiscal Year - fiscal year - interval

Payroll Number - int - ordinal

Agency Name - string- nominal

Last Name - string - nominal

First Name - string - nominal

Middle Init - string - nominal

Agency Start Date - date - interval

Work Location Borough - string - nominal

Title Description - string - nominal

Leave Status as of June 30 - string - nominal

Base Salary - int - interval

Pay Basis - string - nominal

Regular Hours - int - interval

Regular Gross Paid - int - interval

OT Hours - int - interval

Total OT Paid - int - interval

Total Other Pay - int - interval

Evaluation Methods

We can use both visual analysis of figures and correlation as a metric to evaluate trends. Since many of our questions concern change over time, we will be using line graphs. We can use scatter plots for our questions regarding correlation between two variables. These figures will allow us to draw conclusions and compare our results to expected results obtained in previous studies.

Tools

We plan to utilize the online data viewing tool to help with quick filtering and viewing tasks. Cleaning, filtering, and some summary statistics will be done in Prolog. We will also use Python (pandas, numpy, scipystats, matplotlib) to do calculations and generate results.

Milestones

Week 8: Cleaning data, deriving new columns

Week 9: EDA + summary statistics

Week 10: basic visualizations, intermediate progress report -> add to final report

spring break

Week 11: work on questions 1/2

Week 12: work on questions 3/4

Week 13: Put together final project report