New York City Payroll

Nicole Cosmany
Ebrahim Azarisooreh
Stephen Donlin

Project Description

We aim to explore payroll among New York City state employees with respect to borough, department, and time.

Questions:

- How does pay change over time for specific job titles?
- How does salary relate to employment duration? Is this mediated by borough?
- How are total payroll costs distributed by department?
- How much of the city's payrolls costs are allocated to overtime? Are certain departments spending more on overtime?

Datasets

https://data.cityofnewyork.us/City-Government/Citywide-Payroll-Data-Fiscal-Year-/k397-673e/about_data - Citywide Payroll Data - City of New York

Downloaded and available in an interactive online version

Prior Work

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

Overtime spending has been consistently on the rise over the past decade across NYC state employees. Police are generally the biggest overspenders - they had a period from 2015-2019 when OT spending stabilized but has since continued to grow again. In an attempt to mediate overspending, overtime budgets were reduced but this has shown to be ineffective and overtime budgets are continually surpassed. This article focused on 2012-2022 but we will likely find similar trends with our data spanning 2014-2024.

Prior Work

https://www.bls.gov/bls/blswage.htm

The U.S. Bureau of Labor Statistics also, as expected has published datasets and analyses of information pertaining to wage data by occupation over the years. The information provided herein, may support our research or help supplement/replace the datasets provided in this proposal.

Proposed Work

- Look for missing values, check for errors/outliers
- Drop unnecessary columns (dimensionality reduction)
- Consolidate department names to reduce redundancies
- Calculate total payroll and total payroll by department
- Calculate summary statistics - # departments, # employees, payroll ranges, etc.
- Create visualizations - scatter plots, pie charts, histograms

Tools

Python

- Numpy
- Pandas
- Seaborn/matplotlib
- SciPy

Prolog - which has relational/logical features and advanced indexing in the SWI implementation which allows for more concise and efficient querying in the platform of a general purpose tool, which allows for more power and flexibility than SQL.

Interactive government website

- https://data.cityofnewyork.us/City-Government/Citywide-Payroll-Data-Fiscal-Year-/k397-673e/data_preview
- Allows us to download in chunks if needed

Evaluation

To complete this project we will have specific answers to our 4 proposed questions.

We will also have visualizations to support our answers.

We can also make broader connections by searching for additional datasets on job satisfaction by title and agency in an attempt to further analyze relationships in the context of more psychological factors.