Sean Ayoub

Professor Lou

DS 2001

27 October 2022

Project Proposal

1. Solo project.
2. My final project will explore questions such as: Are injured workers fairly compensated? How much is this business paying in overtime costs? Is the system for calculating overtime a liability to the business? Is the overtime system fair? Exploring these questions will help the Human Resources department understand fluctuations in these two categories through the years. This may help the company better plan for future years, decide to hire more employees or redistribute their resources to further support injured workers. Using computational methods will be necessary to sift through large data sets and represent the findings in an easily understood format.
3. For this project, I plan to use the Employee Earnings Report for the City of Boston. This data report consists of a person’s department and title, any injury costs, the amount made in overtime, and their gross salary for the year, among others. This can be used to show trends as well as where the City of Boston may need to make improvements. The data is clean and has been formatted in a CSV file by [data.boston.gov.](https://data.boston.gov/dataset/employee-earnings-report)
4. For this project, I will be grouping items in the data by their department. From here, it will be possible to calculate each different average: the program could display the results for all departments or a specified department of interest. This technique will help answer the questions because it will reveal departments that tend to have more injuries, for example. In this case, the City of Boston could decide to reallocate funds in the injury account in anticipation of the upcoming year.
5. I would like to use bar charts and a boxplot to depict the average salary (and distribution) organized by the department. Furthermore, I can use histograms to show trends in how overtime costs and injured costs have changed throughout the years. These representations will visualize how the City’s current resources are being allocated.

Potential datasets:

[Dataset for People Analytics](https://www.kaggle.com/datasets/koluit/human-resource-data-set-the-company)

[City of Cincinnati Employees](https://data.cincinnati-oh.gov/Efficient-Service-Delivery/City-of-Cincinnati-Employees-w-Salaries/wmj4-ygbf/data)

[Earnings of Female/Male Employees](https://www.kaggle.com/datasets/mpwolke/cusersmarildownloadsearningcsv)

[Retail Store Employees](https://www.kaggle.com/datasets/pmenshih/statistical-data-mbti-of-33k-retail-salespeople)

\*\*\*\* [Employees’ Payroll in Los Angeles](https://www.kaggle.com/datasets/dsfelix/employees-payroll-in-los-angeles)

[Employee Earnings Report - City of Boston](https://data.boston.gov/dataset/employee-earnings-report)