**CS 418: Impact of Covid on Job Growth/Decline Final Report**

**Group 8 - Saahil Sorakayala, Saahi Arumilli, Brian Li, Michelle Zhou, Dawid Biel**

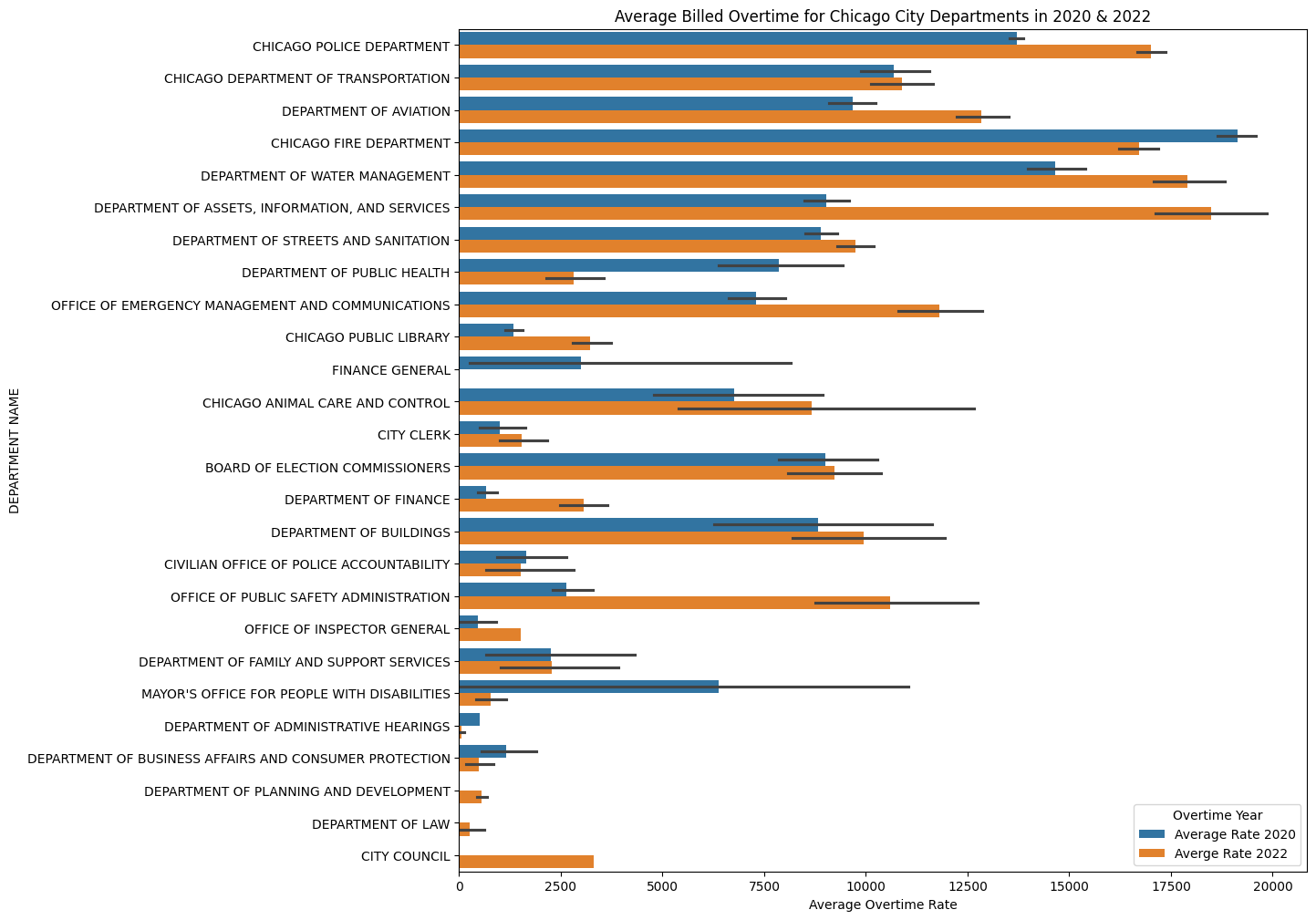
**The Topic:** Throughout this semester we focused on finding out the impact of Covid-19 on job growth and decline in all sectors with a primary focus on the United States. In order to find the answers to this question we looked towards cities like Chicago, states like California and Florida and also national data like those from the Bureau of Labor Statistics. We also went on to expand our scope by looking at data before the pandemic time period to make ML predictions on future outcomes. Throughout this report we will reference our data files which can be found through our github link (<https://github.com/djbiel2/418Final>) and through our shared google drive (<https://drive.google.com/drive/folders/1vHUQNFw7whQPPjPmJSEVvPrI2ThZlcZ3?usp=sharing> ). We highly recommend you view the shared google drive as it has a better log of individual contributions made throughout the semester.

**Visualization 1: Average Billed Overtime in Chicago in 2020 VS 2022**

*Group Member Responsible:* Saahi Arumilli

**The Data:** This visualization was made using data I pulled from the Chicago Data Portal. Specifically the data sets Employee Overtime and Supplemental Earnings 2020 & 2022. To see the final code for this visualization please refer to the github link or google drive link and navigate to the Group 8 Progress Report notebook or Saahi’sProjectDataExploration.ipbn to see all the steps it took to generate this visualization. The data files used to create this visualization are *Employee\_Overtime\_and\_Supplemental\_Earnings\_2022\_20240223.csv and Employee\_Overtime\_and\_Supplemental\_Earnings\_2020\_20240223.csv* please refer to the google drive folder.

**The Results:** From this visualization we saw that there were 7 departments in the city of Chicago that billed more overtime on average during the pandemic than post pandemic. This visualization was a great way for us to kick off looking at what sectors perform better during the pandemic. Specifically, we found that departments such as the fire department, department of public health, and even the department of business affairs and consumer protection performed better during the pandemic than post pandemic.

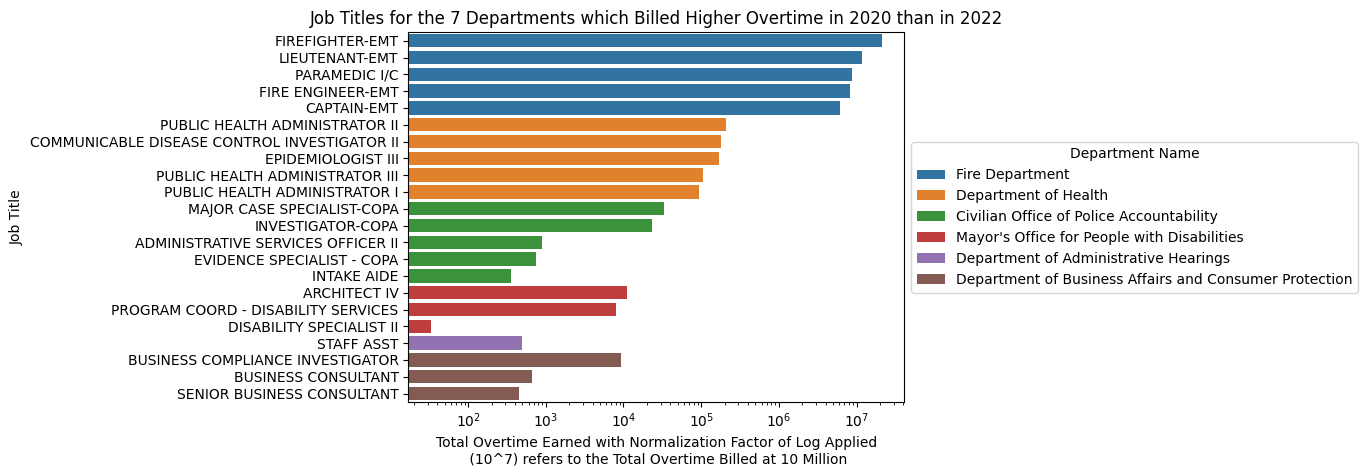


**Visualization 2: The Top Performing Jobs That Billed Higher OT in 2020 in Chicago**

*Group Member Responsible:* Saahi Arumilli

**The Data:** This visualization used the same data sets as Visualization 1 however it specifically looked at the job title column to pinpoint which jobs were attributing to their department performing well during the pandemic. To keep the visualization specific I ended up narrowing it down to only the top 5 contributors. Once again to see the full code please navigate to the group 8 progress report or Saahi’sProjectDataExploration notebook. The data files used to create this visualization are *Employee\_Overtime\_and\_Supplemental\_Earnings\_2022\_20240223.csv and Employee\_Overtime\_and\_Supplemental\_Earnings\_2020\_20240223.csv* please refer to the google drive folder.

**The Results:** The visualization below showed us some foreseeable results such as EMT’s, paramedics, and other CDC and public health officials making more during the pandemic. However, we also saw some not so obvious results in that investigators, program coordinators, and even business consultants were performing very well during the pandemic in Chicago.

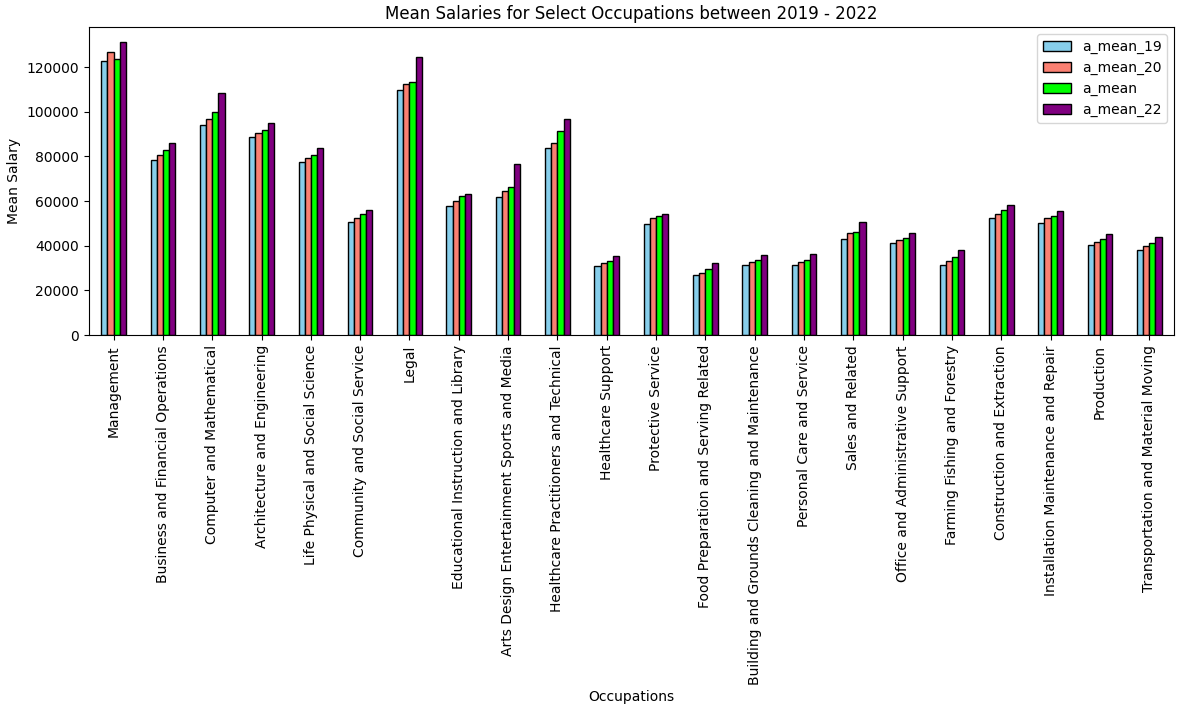


**Visualization 3:**

*Group Member Responsible:* Brian Li

**The Data:** These visualizations look at the annual median and mean salary from 2019 - 2022 by occupation title. The data can be found in our ‘Final Files’ folder, with the name “national\_m20XX\_dl.csv” (replacing XX with the last 2 digits of the year). In order to clean the data, I removed commas, blank entries, and converted values to floats. Lastly, I displayed the values onto a bar chart grouped by occupation. There is one more chart in the notebook not shown here that displays the median.

**The Results:** Overall we noticed a pattern on the mean chart, with all groups having an increase between all years. This could either mean that people of higher pay were being sought after, and/or people of lower pay were being let go. However, in the median chart, all groups saw an increase from 2019 to 2020, but afterwards there was no clear pattern. It can be seen as companies steadying their workforce and rehiring from both sides of the pay range, or hiring from the middle of the pay range.

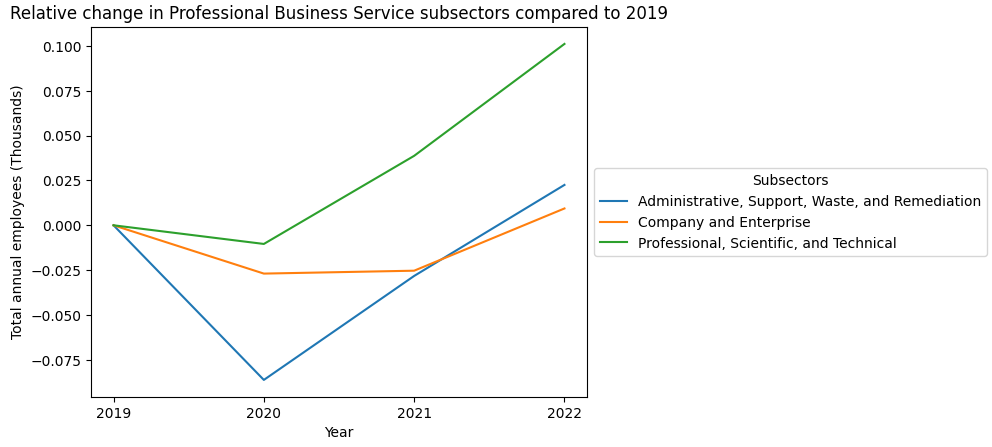


**Visualization 4:**

*Group Member Responsible:* Brian Li

**The Data:** After looking at the growth/decline of different supersectors, we selected 6 that stood out. Reasons being they showed substantial growth, substantial decline, or simply stayed even. In our jupyter notebook, I expanded the supersectors into their sub sectors to gain a deeper understanding of the factors behind the change. The data can be found in our github by navigating into the “Final Files'' folder. Once there, folders for the 6 supersectors can be seen. The code can be found inside the FinalReport.ipynb file. Displayed in this document is the Professional Business Services chart, and the other 5 can be seen in the notebook.

**The Results:** The most drastic growth came from 2 subsectors, which saw a 10% growth from 2019 to 2022. Below is a graph of the sectors that make up the Professional Business Services supersector. During the time of a pandemic, we see that professional, scientific, and technical services skyrocketed, likely from the technical field. Our one concern is whether this would happen again since many companies have already laid the foundation of a remote worklife. Another sector that saw drastic growth was “Building materials, garden equipment, and supplies” in the Retail trade supersector. Lastly, a supersector that failed to recover was “Leisure and Hospitality”, mainly made up by work related to accommodations, recreation, and performing arts.

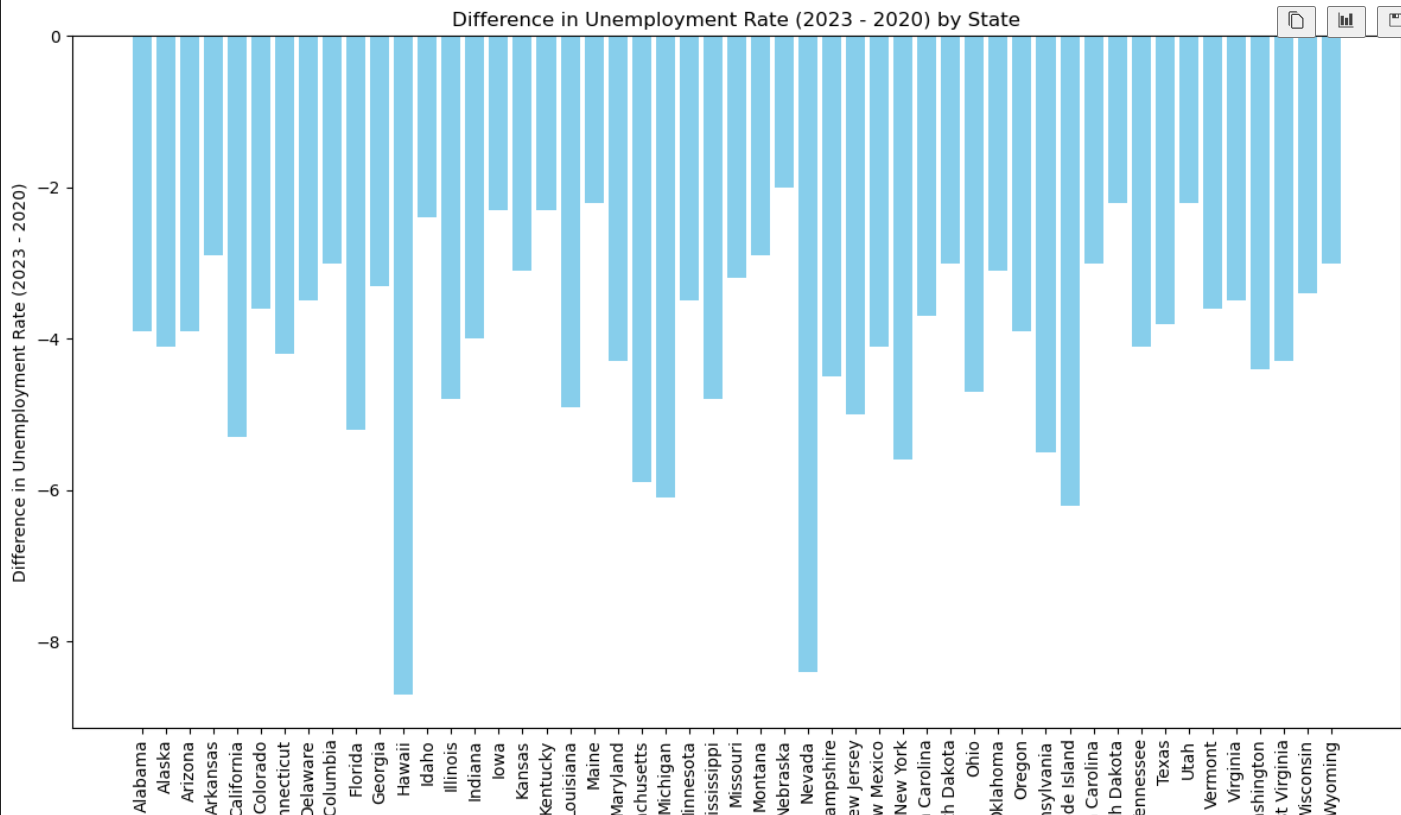


**Visualization 5:**

*Group Member Responsible:* Saahil Sorakayala

**The Data:** Data was cleaned and taken from BLS.

**The Results:** This bar chart shows the unemployment rate of each of the 50 US states. The y-axis represents the differential between 2021 and 2023. These represent the start of the covid pandemic and the end of the covid pandemic. The conclusion that can be made from these results is that all states experienced a lower rate of unemployment by the end of the pandemic. This means that more people became employed by 2023, the end of the pandemic.

****

**Visualization 6:**

*Group Member Responsible:* Saahil Sorakayala

**The Data:**

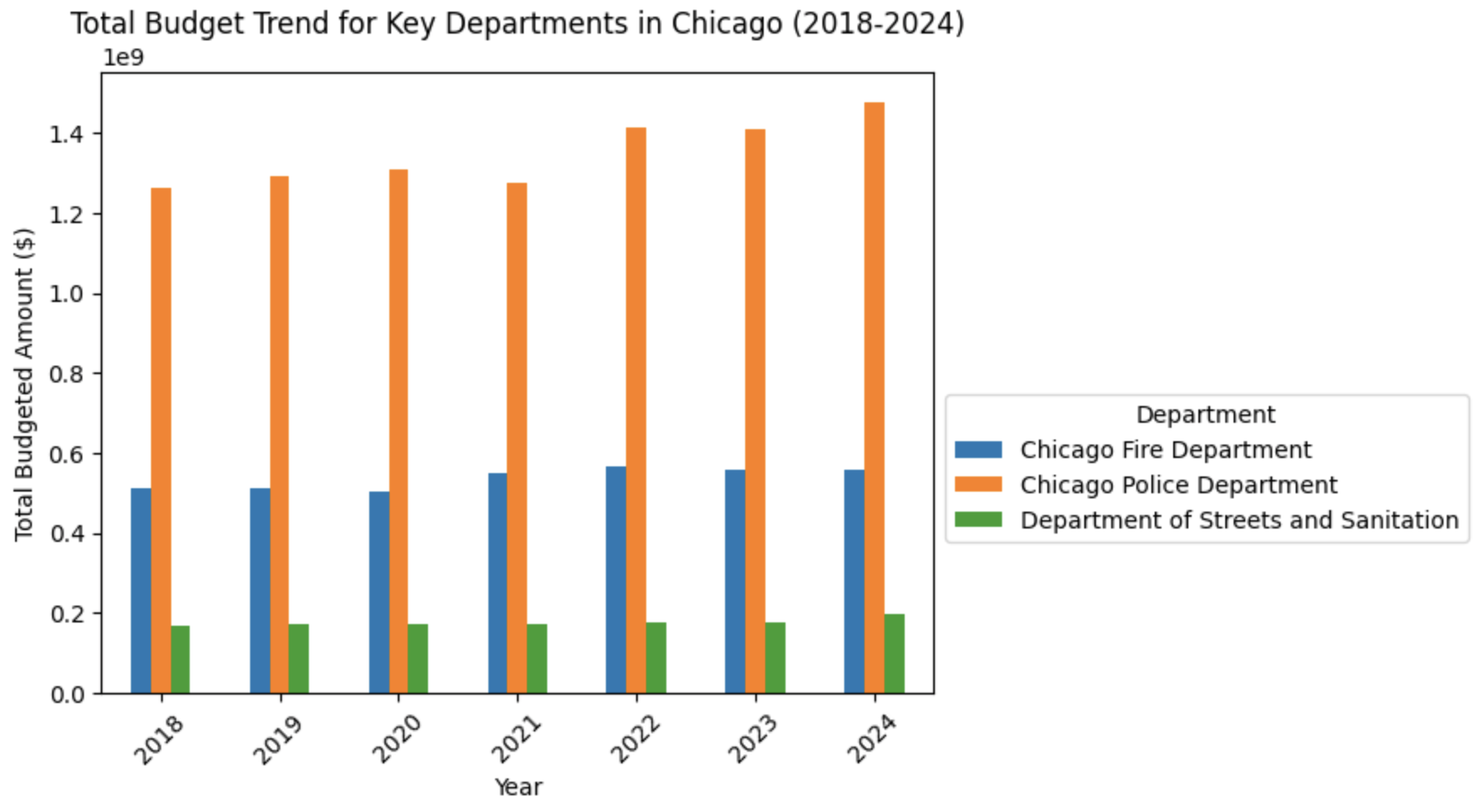
**The Results:**

**Visualization 7:**

*Group Member Responsible:* Michelle Zhou

**The Data:** These visualizations look at the budget allocations to several departments in Chicago that we felt would be impacted significantly by the pandemic. By focusing on departments that were directly involved in critical city functions (such as law enforcement, emergency management, and city upkeep), we hoped to gain information on how budget priorities developed and changed during and following the pandemic. Each of these departments plays a crucial role in maintaining public safety and health, which are profoundly impacted during a public health crisis such as the COVID-19 pandemic.

**The Results:** Though the budget allocation for the Department of Streets and Sanitation remained relatively stagnant, we saw a gradual increase in the budget allocation for the Chicago Police Department for all years except 2021, which is likely due to the fact that the COVID lockdown and social distancing guidelines decreased public disturbances and crime. We see a slight increase in the budget allocation for the Chicago Fire Department in the years following the pandemic, which can be attributed to increased medical emergencies and the need for emergency responses, as well as new protocols that arise as a result of increased guidelines surrounding the pandemic.

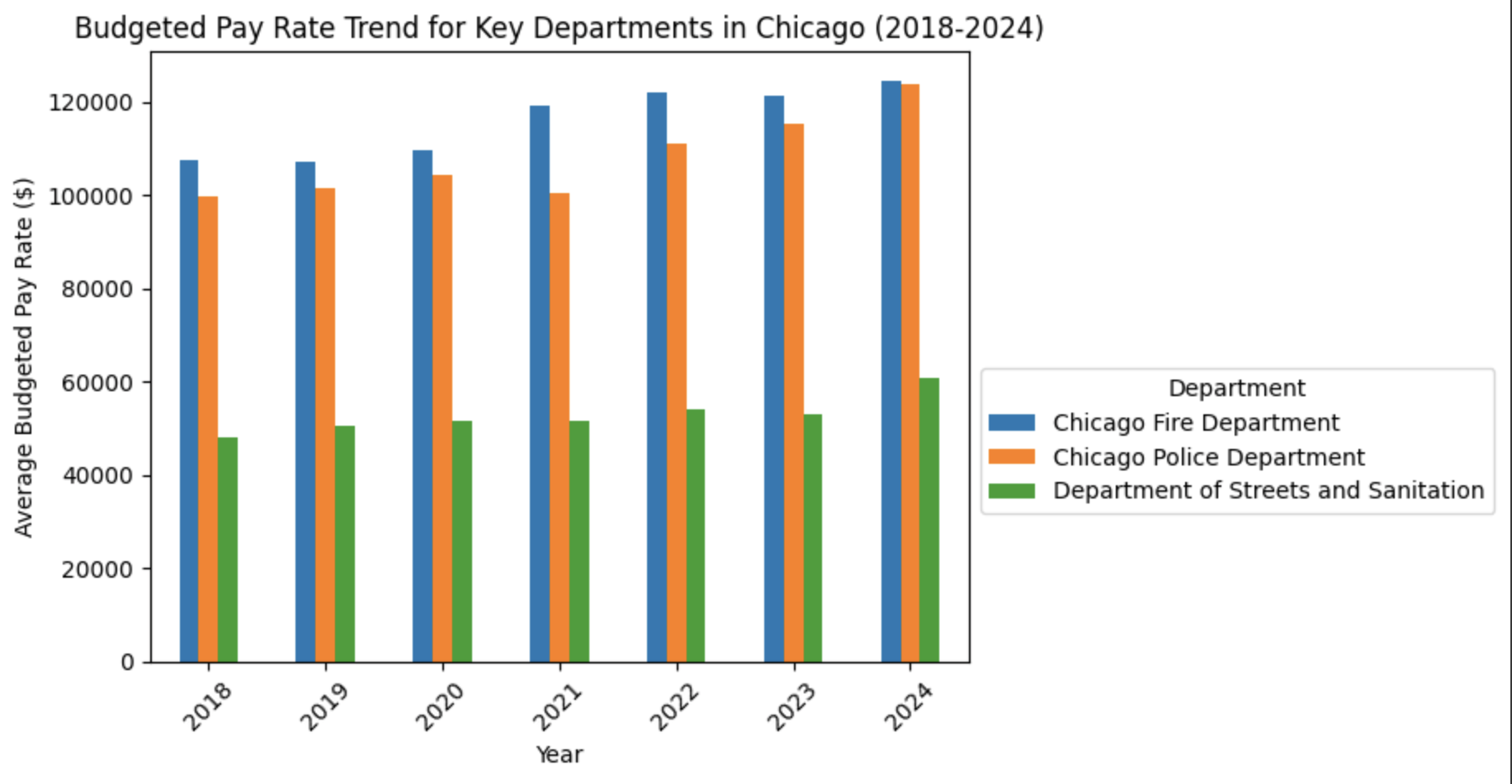
****

**Visualization 8:**

*Group Member Responsible:* Michelle Zhou

**The Data:** Now that we have the results for the total budget trends in the years of 2018-2024, we can compare the budgeted pay rate for the relevant departments to gain insight on how resources are allocated within the respective departments (and which jobs have a higher pay rate). The total budget represents the overall financial resources allocated to each department, covering various expenses such as salaries, operations, equipment, and projects. On the other hand, the budgeted pay rate specifically focuses on the average pay rate allocated for personnel within each department. By comparing the two, we can assess the proportion of the total budget dedicated to employee salaries and benefits.

**The Results:** It seems that, despite the relatively unchanging budget allocation seen from the previous visualization, most departments saw a consistent gradual increase in average budgeted pay rate between 2018 and 2024 except the CPD, which saw a noticeable decrease in pay rate during 2021 (which mirrors the decrease in overall budget allocation in the previous visualization. However, following this dip, the CPD saw a higher rate of growth in average pay rate than the other two departments, ending up with nearly the same average pay rate as the Chicago Fire Department in 2024. The roles and responsibilities within the Fire Department and the Police Department differ significantly, leading to variations in pay scales. During the pandemic, the nature of emergency response work may have changed significantly for both departments. Firefighters and paramedics may have faced increased demands for medical response and public health support, including transporting COVID-19 patients, administering tests, and assisting overwhelmed healthcare systems. Meanwhile, police officers may have been tasked with enforcing public health regulations, managing social distancing measures, and responding to pandemic-related incidents such as lockdown violations or domestic disturbances. These shifts in demand and workload could impact overtime hours, staffing levels, and ultimately, average pay rates. Firefighters, paramedics, and police officers faced heightened health and safety risks during the pandemic, leading to changes in working conditions, protocols, and compensation.



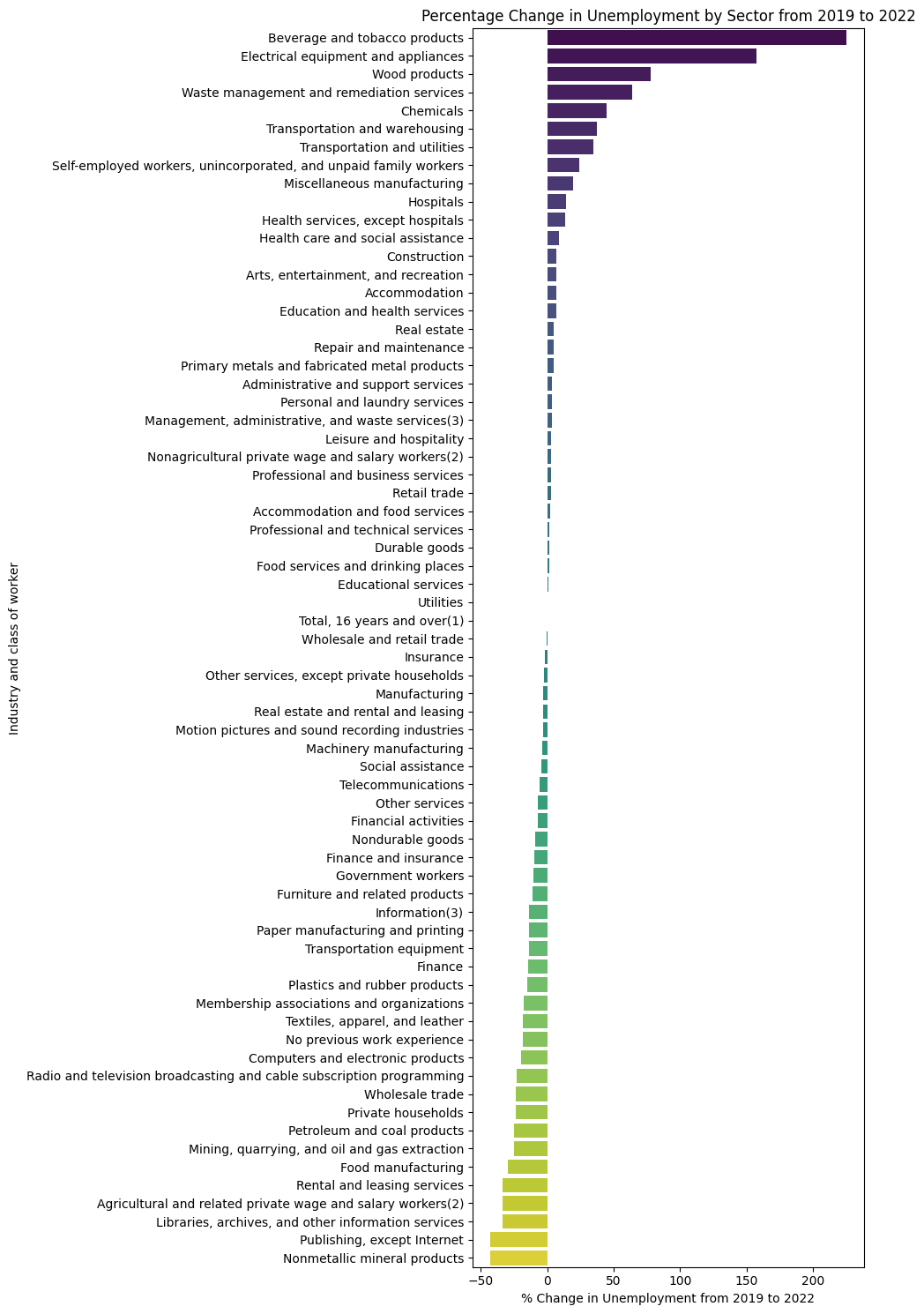
**Visualization 9:**

*Group Member Responsible:* Dawid Biel

**The Data:** Unemployed persons by industry, class of worker, and sex countrywide. Taken from <https://www.bls.gov/cps/cps_aa2019.htm>. I downloaded 4 csv files from the Bureau of Labor Statistics, one for each year from 2019 through 2022 so that we could have a visual representation of the changes in unemployment. The files needed to be cleaned up since each file also had the previous year's data. I made sure that each row's description matched so that I can simply add more columns and add the data to one spreadsheet. It all matched and I successfully copied it into one. The importance of this data is that it shows the trend over 5 years in each industry and further broken down by class of worker and sex. The visualization that I focused on was the overall unemployment change from 2019 to 2022 based on different industries. I did that to fit our parameters. I wanted to take a year prior to the Pandemic and compare it to 2022.

**The Results:**

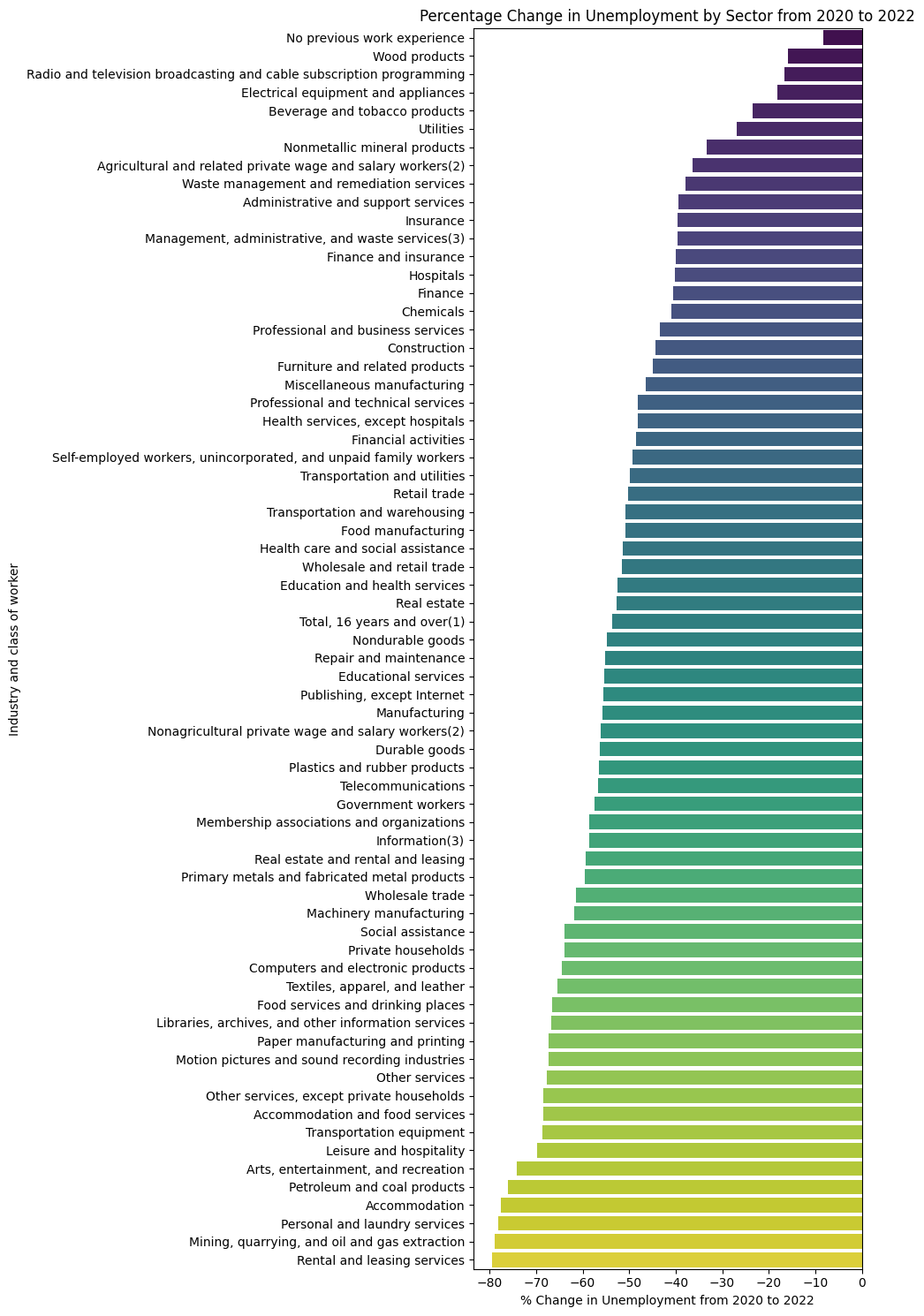
The first visualization is from the period 2019 to 2022. It shows the percentage change in Unemployment by sector. It shows a very clear indication of which industries had a positive change and a negative change in unemployment. The general trend however is that there are 3 separate categories. Those that had zero to minimal change, and those that did have a drastic positive or negative change. Through that inference I was able to make the best k means cluster for my ML analysis. From the chart we see that the Beverage and tobacco products sector had the biggest positive change in unemployment while non metallic mineral products had the biggest negative change. We can use this visualization to infer for the future which industries will have a higher than normal change in unemployment. Which sectors will see a boom and which will significantly contract.



**Visualization 10:**

*Group Member Responsible:* Dawid Biel

**The Results:** This visualization uses the same data as the one above, but the only parameter that I changed was the start year. This visualization is from 2020 to 2022. I wanted to have both of these to show the sharp contrast between the numbers. This graph specifically references the pandemic. We see that all industries had a negative change in unemployment. We can see that those with no previous work experience had the worst numbers. However it is still important to point out that it was overall a negative change in unemployment. We see that during covid Rental and leasing were affected the most, but there was a sharp rebound in those jobs.



**ML Analysis 1: Predicting the Future Budget of Chicago using Linear Regression**

*Group Member Responsible:* Saahi Arumilli

**The Data:** Given that there are various fluctuations that occur in an economy with the pandemic being one of them I wanted to look at if I could accurately predict the future budget of the city of Chicago. I used linear regression from the sklearn package and pulled my data once again from the Chicago Data Portal. I used the Budget recommendation - positions and salaries data sets from 2012 - 2021 to train my model. I then used the budgets for 2022 and 2023 to check the accuracy of my linear regression predictor. Please refer to the Group 8 Progress report or Saahi’s Project Data Exploration.ipbn to see the full code used to make this visualization. Please also see the data files titled 20(12-23)ChicagoBudget.csv in the shared google drive folder under the final files folder and then under the *Saahi-Data-Files* folder to see the full data.

**The Results:** LinearRegression(). Linear Regression Coefficient Score: 0.9674431578517175. This suggests that there is a strong positive correlation as the years keep going the overall budget needed to run the city is increasing. From the below values we can see that our linear regression prediction model does under-report the budget for the predicted years but they are only off by about a hundred million or so.

The Predicted Budget for 2022: $3,281,116,314.53 - The Budget for 2022: $3,393,276,423.0

The Predicted Budget for 2023: $3,352,232,500.85 - The Budget for 2023: $3,522,920,659.0

In the future as more data becomes available this model can improve and make better estimations on how to plan the budget of the city. This can also greatly advise on the overall cost per employee metric the city would have to consider in making future budgets. While our question focuses on the impact of Covid, this model provides us with a great way to compare the real budget issued, the estimated budget, and can even take into account any emergency funds provided in the time of an emergency.

**ML Analysis 2:**

*Group Member Responsible:* Brian Li

**The Data:** Here I wanted to see if there was a correlation between the salary and the people that were let go. The best way I could think was using the percentile salaries, and trying to make a model predict the median salary. Using the 2019 data for training and the 2020 data for testing, I performed a ridge regression using multiple features. The data files are named “national\_M20xx\_dl.csv” in ‘Final Files’. There was not much data manipulation besides cleaning values to be converted into floats.

**The Results:** After switching the features to be used, the lowest MSE my model could obtain was 884916.67, much better than my initial 2+ million. From this however, I came to the conclusion that there was not as much correlation as I had expected. The features used to obtain the lowest MSE were the 10th, 25th, 75th annual percentile salaries, and the mean. Initially I also had the 90th percentile salaries, but that gave a high error because there were some empty values. I also tried swapping the median to be used as one of the features, and attempting to predict any of the percentile wages. But I reverted to my original features as this change resulted in an even higher error.

**ML Analysis 3:**

*Group Member Responsible:* Saahil Sorakayala

**The Data:** For this analysis the data for Visualization 1 was analyzed to predict the changes in overtime billing between 2020 and 2022. There were 3 models, (linear regression, logistic regression, and SVC) used in order to compare which model would be suitable for predicting an increase or a decrease in overtime billing. I used the same data from the Chicago Data Portal. Then I calculated the total for each department, merged the data into one dataframe, and created the features based on the difference between the totals, and the labels based on if the difference between the totals was positive or negative. These were input into the 3 models, and then compared the results.

**The Results:** We found that the model provides a reasonable estimate of overtime costs for most departments. However, there may be some departments where the model performs better or worse than average. For instance, departments with relatively stable overtime patterns over the years may have lower prediction errors, while departments with more volatile overtime trends may exhibit higher prediction errors.

-------------------------------------------------------------

Training with Linear Regression Model

Mean Squared Error: 0.2310986924713772

R-squared: 0.03708878136926186

Coefficients: [3.54520935e-08]

Intercept: 0.6181193961534278

-------------------------------------------------------------

Training with Logistic Regression Model

Accuracy: 1.0

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 2

1 1.00 1.00 1.00 3

accuracy 1.00 5

macro avg 1.00 1.00 1.00 5

weighted avg 1.00 1.00 1.00 5

Coefficients: [0.0175446]

Intercept: 8.835222524502319e-05

-------------------------------------------------------------

Training with SVC

Accuracy: 0.8

Classification Report:

precision recall f1-score support

0 1.00 0.50 0.67 2

1 0.75 1.00 0.86 3

accuracy 0.80 5

macro avg 0.88 0.75 0.76 5

weighted avg 0.85 0.80 0.78 5

**ML Analysis 4:**

*Group Member Responsible:* Michelle Zhou

**The Data:** For this ML Analysis, I was interested in seeing if there was a way to predict the overtime costs for various Chicago City departments based on the earnings data. I used two main datasets–the Chicago\_City\_Employment\_Data\_20240223.csv and the "Employee\_Overtime\_and\_Supplemental\_Earnings\_2020\_20240223.csv" datasets. I chose to use Linear regression and MSE (Mean Squared Error) to predict and assess the results from the data.

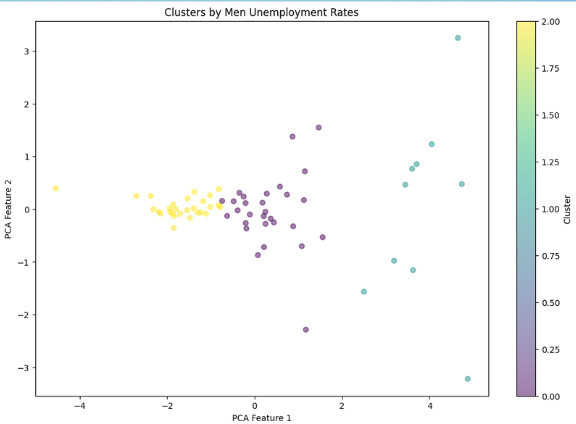
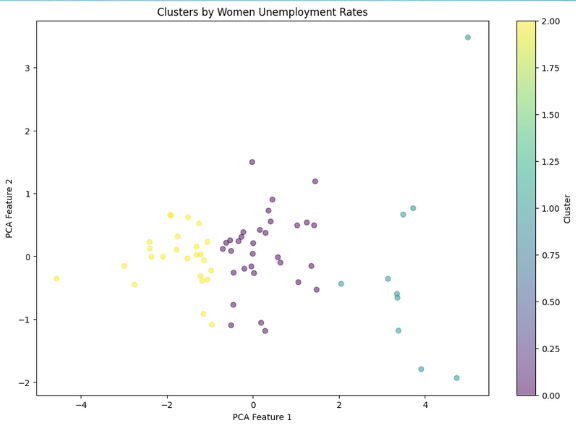
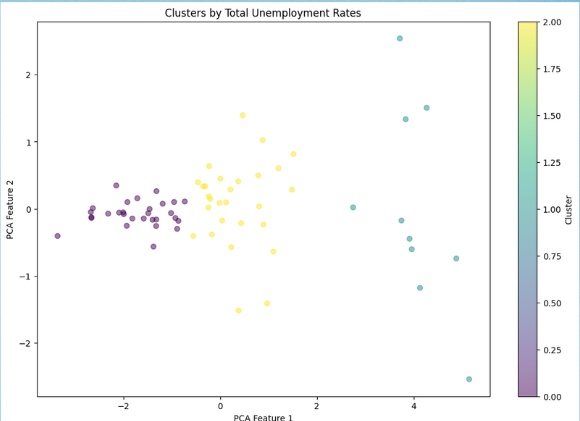
**The Results:**

**ML Analysis 5:**

*Group Member Responsible:* Dawid Biel

**The Data:** For the ML analysis I wanted to focus on the correlation between the changes in overall unemployment based on the sectors. I used the same data, but even further cleaned up to make Cluster analysis easier. The data that I collected had a breakdown between men and women in terms of Unemployment with respect to each sector. I wanted to analyze if there are any sectors or relationships between those variables.

**The Results:** From the first visualization I chose K means to equal to 3 due to there being three different clusters that I observed. It ended up being the best match for this data. I tried various others but did not observe clustering properties. I have 3 visualizations, one for total and one for each sex. I specifically looked for close clusters. The colors do not match “groups”, but in the code it does print out what each point corresponds to. I found it interesting looking at the difference between men and women. The men’s cluster 2 had a very densely populated cluster while the women’s had 3 clusters but they were still spread out. From the other Visualizations as well as my research into %of men vs women in certain industries it became apparent that some of the industries that men work in had a closer relationship in terms of unemployment change. I believe it is due to the fact that men and women tend to work in different industries, hence there is a big percentage difference of men vs women in them.

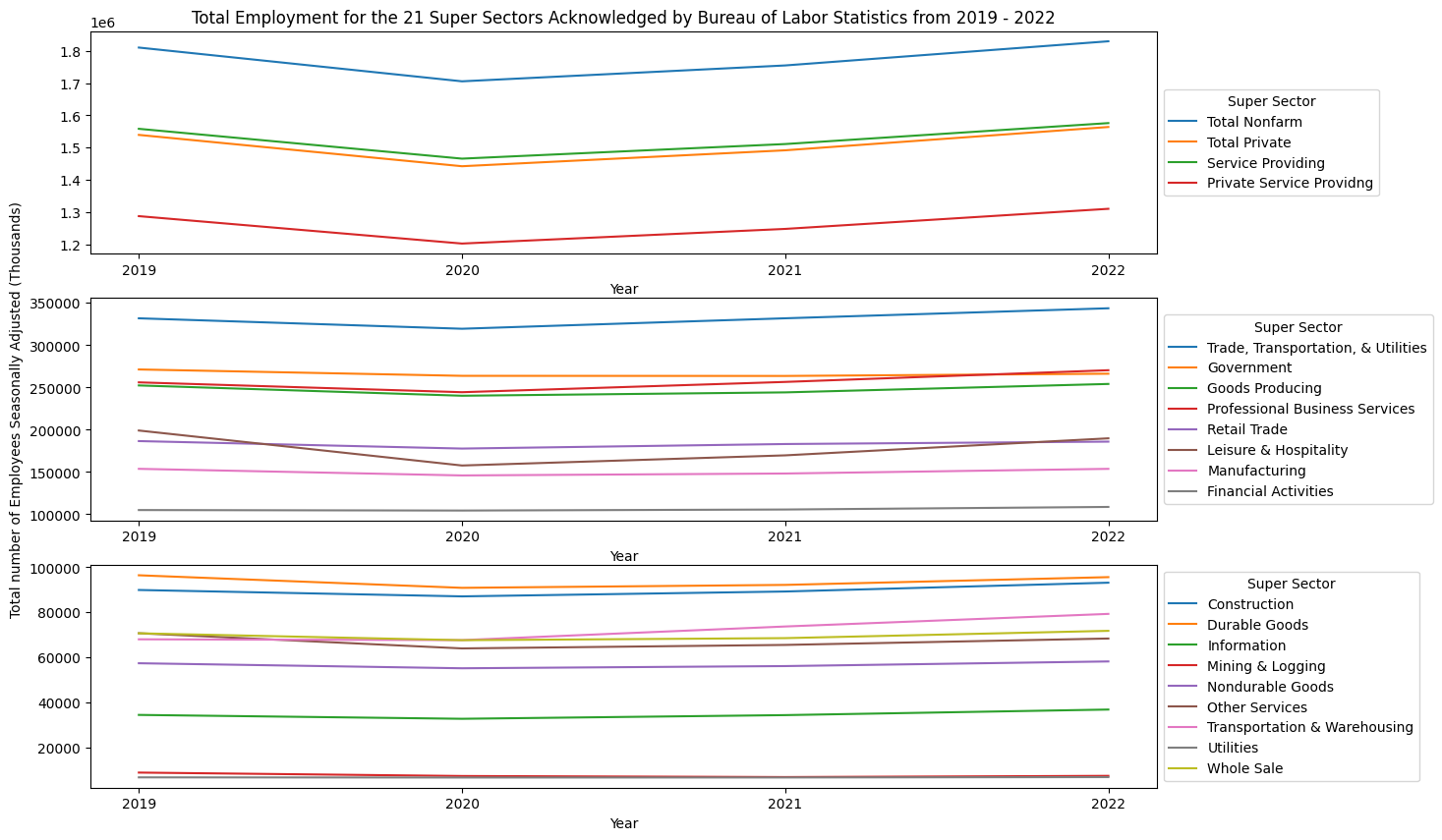


**Additional Work 1: Total Employment for BLS Supersectors from 2019 - 2022**

*Group Member Responsible:* Saahi Arumilli

**The Data:** The following visualization was created using data from the Bureau of Labor Statistics. Specifically, this visualization looks at the Total Employment seasonally adjusted for the 21 super sectors noted by BLS from 2019 - 2022. The data had to be cleaned as there were many header comments and additional I had to create a total column to get the total employment for the entire year as the data was given in months. In total there were 21 xlsx files I had to clean and then convert to csv files to make this visualization. To view the code for this visualization please refer to the Group8\_Progress\_Report.ipbn or Saahi’sProjectDataExploration.ipbn to see all the steps. To see the data please navigate to the google drive folder and the final files subfolder and then to the folder named *Saahi-Data-Files*, there refer to files starting with the “All” tag to see the super sectors used in this visualization. To see the individual graphs created per each supersector please refer to the *Graphs for each super sector.pdf* located in the ‘*Saahi-Data-Files*’ folder. To make this visualization I relied on seaborn line plots as well as splitting the data into three different frames as some sectors had very large gaps between other sectors.

**The Results:** In this visualization we found that every single super sector did in fact have a dip in employment in 2020. The pandemic hampered the employment rate of many supersectors but, there were many sectors that were able to bounce back and have even surpassed their pre pandemic employment levels. Some of the sectors doing better include construction, financial activities, private business services, and trade, transportation, and utilities. Some sectors that have not been able to recover since the pandemic include leisure and hospitality, government, mining & logging, retail trade.

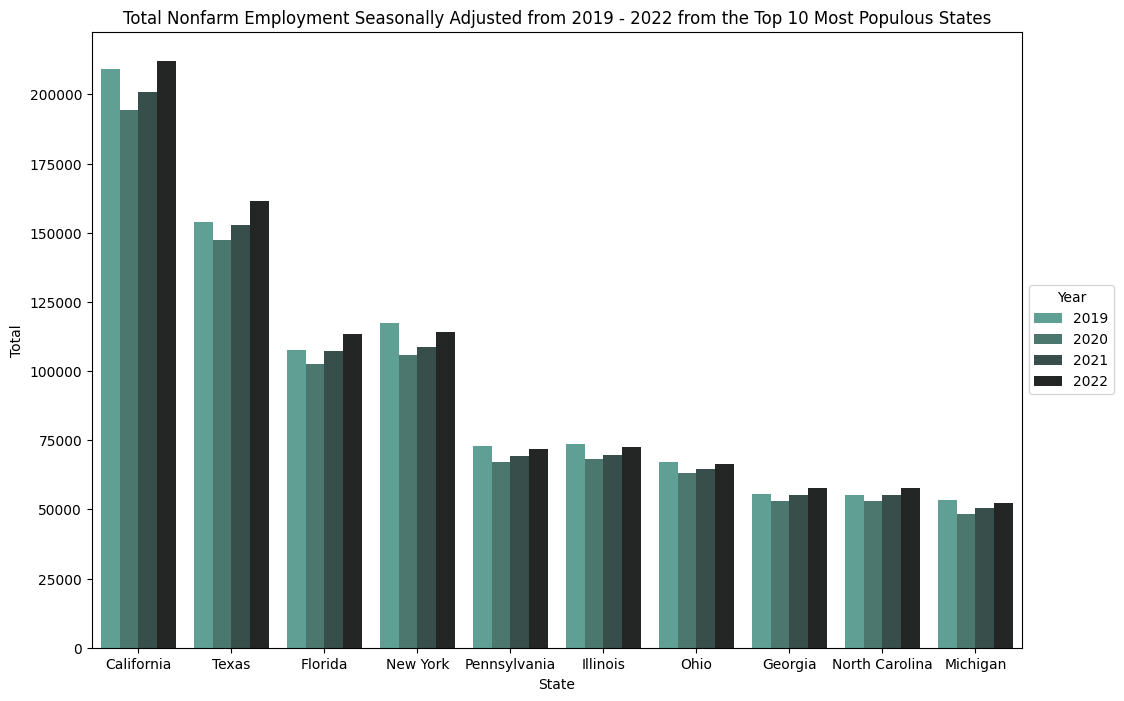


**Additional Work 2: Total Nonfarm Employment in Top 10 Populous US States 2019 - 2022**

*Group Member Responsible:* Saahi Arumilli

**The Data:** This visualization focuses on the top 10 most populous states in the United States. I found the total nonfarm employment with seasonal adjustment from 2019 - 2022 from the Bureau of Labor Statistics. The 10 states are California, Texas, Florida, Pennsylvania, Illinois, Ohio, Georgia, North Carolina, and Michigan. The data files for these states are located in the google drive folder State(state name)Emp.csv. Please refer to the final files folder then the folder name *Saahi-Data-Files* to see the full data. The code for this visualization can be found in both the Group8\_Progress\_Report.ipbn or for the comprehensive code written please see Saahi’s ProjectDataExploration.ipbn. To clean the code I had to remove the header comments that automatically downloaded from the BLS files and I had to create a totals column as the data was split by months. I also changed all the xlsx to csv. After this I used the pd.concat function to join all 10 states together and used a seaborn barplot to show the data.

**The Results:** From this visualization I found that half of the states in the top 10 have just barely hit their pre pandemic levels of employment or have just barely surpassed it. This is quite concerning as these 10 states produce a bulk majority of the US’s revenue. This means that even till 2022 the economy was still in an extremely compromised state.

****

**The Overall Results:** We found that the pandemic impacted all areas of the economy especially when considering the year 2020. Certain industries like that of healthcare were doing significantly better, but this was largely due to the nature of the pandemic and the vast need for healthcare workers, epidemiologists, and others in that field.

Post pandemic we saw that some sectors were particularly fast in rebuilding their industries. In some cases we can even say that many of the sectors now have the infrastructures in place to handle another crisis if it does occur, though we greatly hope that this does not happen.

We would also suggest to those who are not in the healthcare industry or lack skills in that sector to focus on gaining business acumen or even accreditation as we found significant evidence that those with a combination of their industry skills and business knowledge have been able to perform well during periods of economic uncertainty.