COVID Unemployment - Report

Riya Pandey (STAT 612) and Rebecca Andrade (STAT 612)

2023-05-09

Executive Summary:

Our group project aimed to understand the impact of COVID-19 on employment in the US by analyzing data on age, gender, race, and industry. Using publicly available data from the Bureau of Labor Statistics, we performed descriptive analysis and statistical tests to identify trends and patterns. Our primary questions of interest were whether the pandemic caused a disparity in employment for men and women, how age and race affected employment, and which industries were most and least affected.

Our analysis revealed that women were more affected by job losses than men. Younger workers and Asian and Black/African American races had higher job loss rates. The most affected industries were leisure and hospitality and retail trade.

Overall, our findings highlight the significant impact of COVID-19 on employment in the US and the need for targeted policies to address the disproportionate impact on certain demographic groups and industries.

Introduction

The COVID-19 pandemic has had a significant impact on the global economy, leading to widespread job losses and economic instability. The United States has not been immune to these effects, with the pandemic causing significant disruption to the employment situation in the country. Our project aimed to analyze the impact of the pandemic on employment in the US, with a focus on understanding the disproportionate impact on certain demographic groups and industries. Specifically, we were interested in examining whether women, young workers, and workers of African American and Hispanic ethnicity were disproportionately affected by the pandemic in terms of employment, and which industries were most affected. By understanding these impacts, we hoped to provide insights that could inform policy decisions and support efforts to mitigate the negative effects of the pandemic on employment.

There were several motivations behind the research that we conducted. Specifically, we sought to understand the economic and social impacts of the COVID-19 pandemic. It was important for us to see which industry sectors were being hit the hardest because understanding the impact of employment rates aids us in understanding how policy makers make

decisions for support measures. The social implications have also been striking during the pandemic era. Loss of jobs has led to financial strain and understanding how employment losses have affected women, young adults, and minorities may help to address the visible disparities of the impact across demographics.

To conduct our analysis, we used publicly available data from two sources, the Current Population Survey (CPS) and the Current Employment Statistics (CES), both of which are provided by the US Bureau of Labor Statistics (BLS). We performed descriptive analysis and statistical tests to identify trends and patterns, focusing on demographic variables such as age, gender, and race, as well as industry-specific variables. The data set we used included observations from 2015 to 2022 to observe trends and patterns but we focused mainly on the years 2019 and 2020. With a large sample size, we were able to draw reliable conclusions about the impact of COVID-19 on employment across different demographic groups and industries.

Literature Review

A systematic review on the impact of COVID-19 on employment rates, has proved very informative in conducting our analysis. Previous studies have documented the significant impact of the pandemic on employment, with some studies highlighting the disproportionate impact on certain demographic groups and industries. A study by the Economic Policy Institute found that women, young workers, and workers in low-wage industries were more likely to experience job losses during the pandemic (Economic Policy Institute, 2020). Other studies have highlighted that women have been more affected by job losses than men during the pandemic Center for American Progress, 2021). Similarly, there is evidence to suggest that younger workers and minority groups have experienced higher rates of job loss ("Racial Disparities in COVID-19, 2020). Looking specifically at minority groups that have been large disparities in job loss particularly in the Asian American community, which has been hit particularly hard by the crisis (Springer Link, 2022).

These studies highlight the importance of examining the differential impact of the pandemic on employment in the US. Reviewing several sources and conducting our analysis has allowed for a greater understanding of the correlation between different demographic groups and job loss rates, particularly in 2019 and 2020.

Initial Hypothesis

1. Women, young workers, and workers of African American and Hispanic ethnicity are disproportionately affected by the pandemic in terms of employment compared to other demographic groups.

2. The pandemic has varying impacts on different industries, with the Leisure and Hospitality Industries and transportation being most affected (in other words, is in the "Affected" group) and private education and health services being least affected (in other words, is not in the "Affected" group).

Data Preparation

Data preparation is always a critical step in any data analysis project. We encountered a few challenges in gathering and tidying the data, but we used a variety of approaches to overcome them.

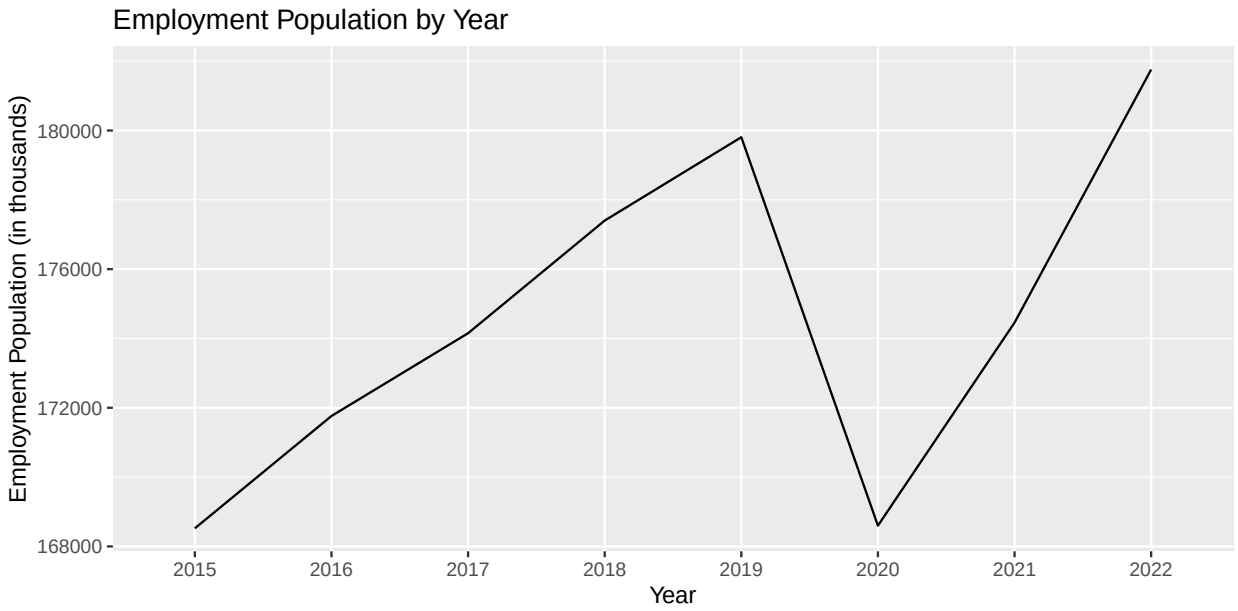
One of the main challenges was identifying and accessing reliable data sources that provided detailed information on employment trends during the pandemic. We used data from multiple sources from the Bureau of Labor Statistics, including the Current Population Survey (CPS) and Current Employment Statistics (CES) program. Since the data was from a government agency, we had minimal complications.

The quality of data was very important and another challenge we encountered was the inconsistency of the data, which included missing values, errors, outliers, and inconsistencies in coding and it took us a while to find that. To address this challenge, we cleaned and tidied the data by removing any missing values, correcting errors in the coding, and ensuring that all the variables were in correct format for our analysis.

We encountered some challenges related to cleaning and tidying the data and transforming the data to create a cohesive data set that could be analyzed in a meaningful way. To address this challenge, we used 'tidyr' and 'dplyr' to merge and reshape the data as needed.

Exploratory Data Anlaysis

Employment Population by Year



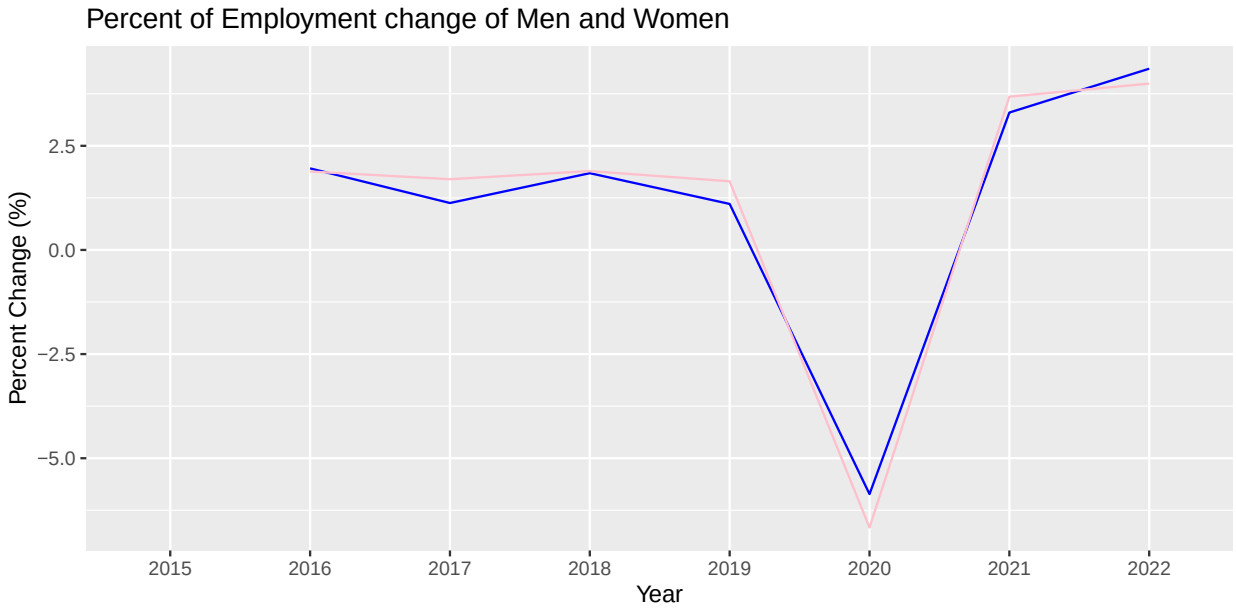
In our exploratory data analysis, we first wanted to get an overview of the employment population to look at the trends over time from 2015 to 2022. We created a line graph to visualize these trends, which showed that the employment population remained relatively

stable from 2015 to 2019, with an upward trend. However, we can see a significant drop in the employment population from 2019 to 2020, which were the years of the COVID-19 pandemic.

OLS regression for initial analysis

[Appendix 1] After conducting initial exploratory data analysis on employment population and year, we conducted an initial OLS regression to investigate the impact of demographic variables such as age, gender, and race on employment population during the pandemic. The regression results indicated that all the demographic variables had a significant impact on employment status during the pandemic. The variable Year may not have shown significance in the model because the impact of the pandemic might not have been significant enough in the earlier years to have a noticeable effect on the overall employment population. Therefore, the inclusion of these earlier years might have diluted the effect of the pandemic on the employment population and reduced the significance of the variable Year in the model.

Percent Employment Change of Men and Women in 2019 and 2020



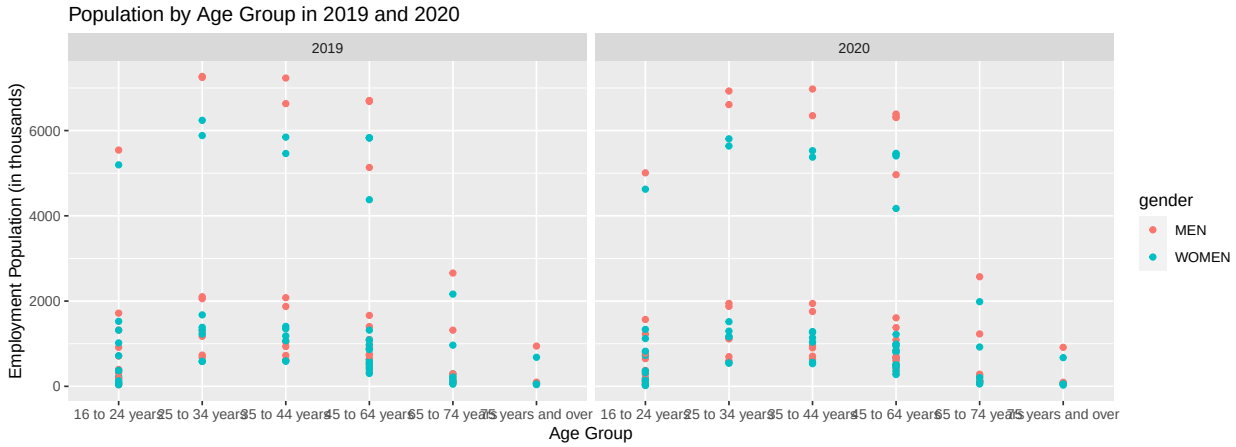
To explore the impact of the pandemic on employment between genders, we calculated the percentage change in employment for each gender for all the years from 2015 to 2022. We then created a line graph to compare the data. Our analysis showed that there was a similar trend in employment between men and women from 2015 to 2019, with a slight increase over the years. However, there was a significant drop in employment for both genders in 2020 due to the pandemic. The percentage change in employment from 2019 to 2020 was -6.66% for

women and -5.86% for men, which indicates that women experienced slightly more job losses than men during the pandemic. This is supported by the line graph, which shows a sharper dip in the percentage change in employment for women compared to men in 2020. This finding highlights the disproportionate impact of the pandemic on women's employment and the need for gender-sensitive policies and initiatives to support women in the workforce.

Statistical Analysis: t-test

To confirm our hypothesis, we conducted a t-test to compare the mean employment population of men and women before and after the COVID-19 pandemic. To perform the analysis, we first calculated the mean employment for both men and women for the years 2015-2019 (before COVID) and for the years 2020-2022 (after COVID). The t-test *[Appendix 2]* is a statistical test used to compare the means of two groups. In our case, we compared the mean employment numbers of men and women before and after COVID-19 using the t-test. The results showed a significant difference between the mean employment numbers of men and women. Men had a higher mean employment than women, indicating that women were more disproportionately affected by the pandemic in terms of employment compared to men. The p-value of 0.0005347 is less than the significance level of 0.05, which means we can reject the null hypothesis that the mean employment for men and women is the same. Additionally, the confidence interval of 95% shows that the true difference in means is likely between 216.88 and 254.12. This suggests that men experienced a smaller decrease in employment population compared to women during the COVID-19 pandemic. Therefore, our results indicate that women were indeed disproportionately affected by the pandemic in terms of employment, confirming our initial hypothesis.

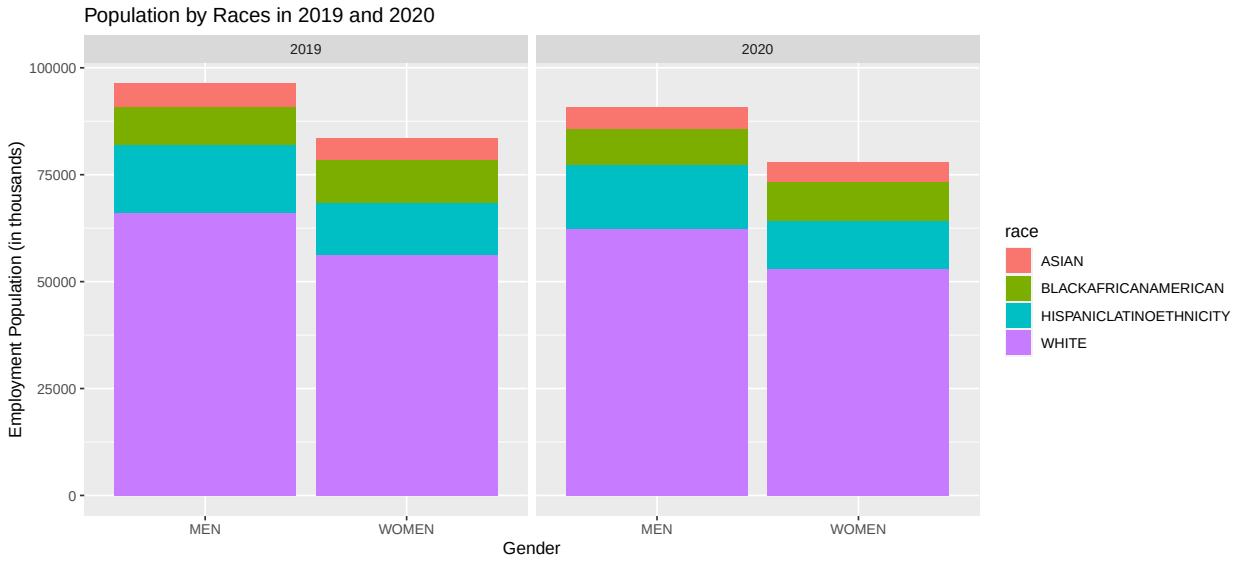
Employment Population in different Age Groups filtered by gender (2019 and 2020)



After our initial analysis, we wanted to delve deeper into the impact of COVID-19 on employment. Therefore, we performed an exploratory data analysis on the data for the years

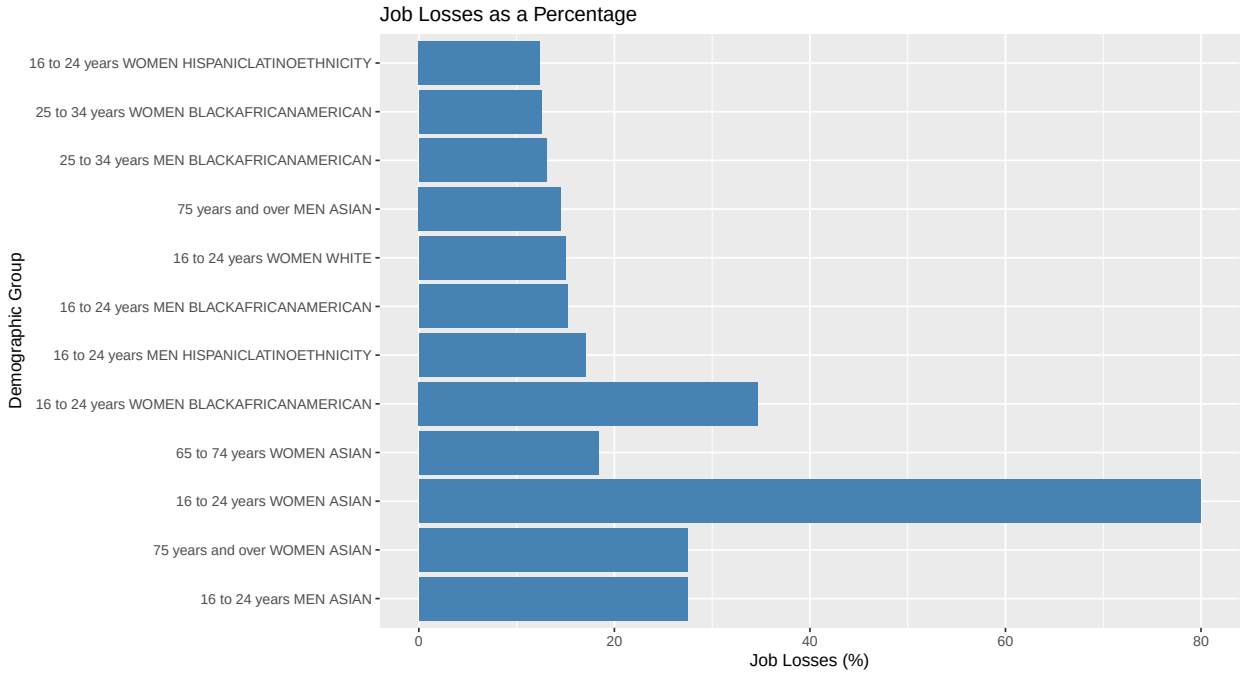
2019 and 2020, focusing on other variables. To start, we created a scatter plot to show the employment population in different age groups, filtered by gender, for these two years. The scatter plot illustrated that the age groups of 25-34 and 35-54 had the largest employment population for both men and women during these years. We also observed a significant drop in the employment population across all age groups between the two years. This decrease was especially notable for those aged 16-24. These findings suggest that the COVID-19 pandemic had a significant impact on the employment population across all age groups, but especially for older individuals.

Employment Population in different races filtered by gender (2019 and 2020)



After exploring the data for other variables, we wanted to investigate the impact of race on employment during the pandemic. To do this, we created a stacked bar chart to compare the employment population of different races, filtered by gender, for the years 2019 and 2020. The chart revealed that the white race had the highest employment population in both years, followed by Hispanic, Black, and Asian. We observed a decrease in employment population across all races from 2019 to 2020. However, we were not able to identify which race had the most significant decrease from the chart alone. Therefore, we conducted further analysis to gain a better understanding of the impact of race on employment during the pandemic.

Impact of the pandemic on younger workers across different races and gender



To gain a better understanding of the extent of the pandemic's impact on employment, we conducted an analysis on the percentage change in employment across various demographic variables. This included race, gender, and age. We then created a bar chart to visualize the top demographic groups that were most affected. Our findings revealed that the top three demographic groups with the highest percentage of job losses were predominantly women. Specifically, the group that was most impacted were women of Asian ethnicity between the ages of 16 and 24, with a 44% decrease in employment. This analysis confirms that women of age group of 16 and 24 of Asian ethnicities were impacted the most during covid. This analysis confirms our hypothesis that the younger workers were more impacted than older workers and minor ethnicities, especially Asian races, were more impacted than other races.

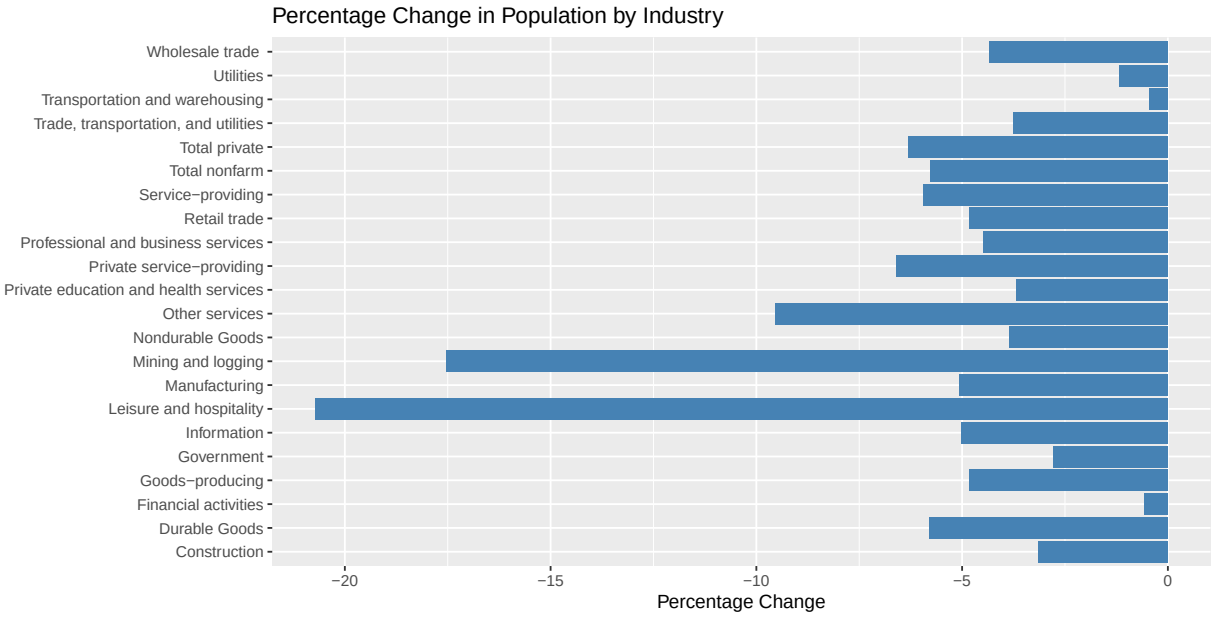
Staistical Analysis: ANOVA and Linear Regression

Our model was built to explore the relationship between population size and age and race variables. The intercept term of the model predicts the population size when all other variables are zero, and in this case, it's 5053. The coefficients for Age_Group and race variables demonstrate how much the predicted population size changes for each level of each variable relative to the reference level. *[Appendix 3]* Our analysis shows that all age groups, except for 25 to 34 years, have a significant impact on the population size. People aged 16 to 24 and 75 years and over have the most significant effect on population size, with coefficients of -1472.2 and -2112.2, respectively. Similarly, all race variables significantly impact population size, with people of Asian ethnicity having the most significant effect,

followed by African Americans and Hispanic/Latino ethnicities. The ANOVA table results show that both Age_Group and race are significant predictors of population size, and the R-squared value of 0.7714 indicates that 77.14% of the variance in population size can be explained by the Age_Group and race variables.

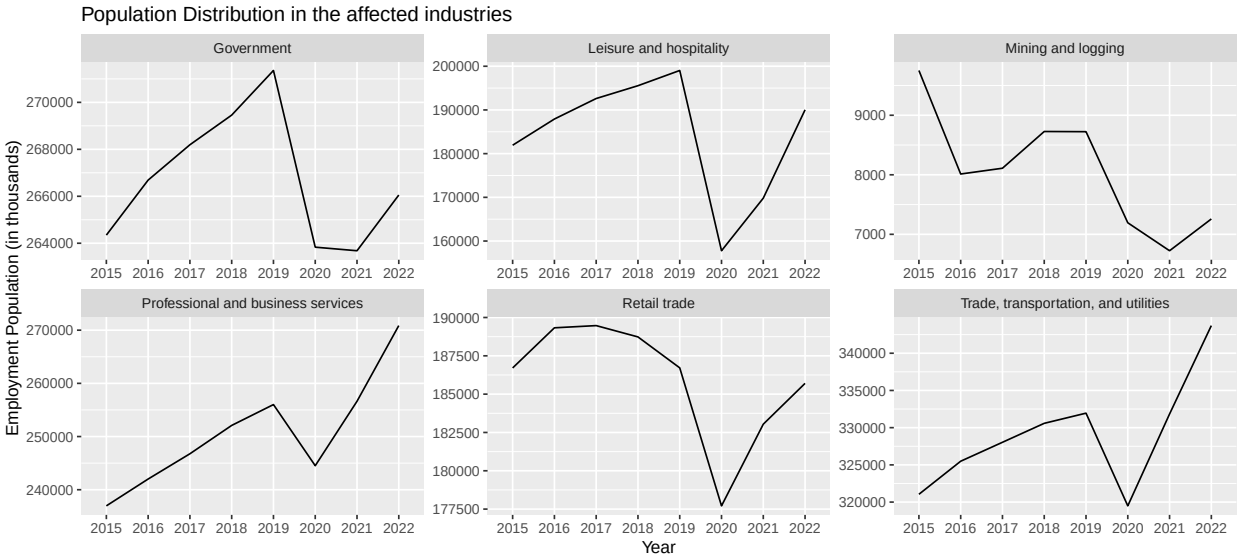
Industry

Bar chart of percentage change in population for each industry from 2019 to 2020



After analyzing the data on industries, we calculated the percentage change in employment for each industry between 2019 and 2020. To better visualize this information, we created a bar chart. The results showed that leisure and hospitality had the highest percent decrease in employment population, followed by mining and other services. With our knowledge of industries, we then created two groups - Covid_affected and Covid_unaffected - and assigned different industries to each group based on their performance during the pandemic. The Covid_affected group included industries that were hit the hardest by the pandemic, such as leisure and hospitality, while the Covid_unaffected group included industries that were relatively less affected, such as healthcare and finance. This grouping allowed for a more focused analysis of the impact of the pandemic on different industries and provided insights into the varying levels of vulnerability among different sectors.

Affected Supersectors EDA



In order to understand the impact of the COVID-19 pandemic on various industries, using the bar graphs and our knowledge on industries affected during covid, we categorized them into two groups: "COVID affected" and "COVID unaffected". The classification coincided with the employment magnitude of the supersectors, with the "Affected" group consisting of the five largest supersectors by total employment, each with over ten million employees, and the "Unaffected" group consisting of the remaining supersectors, each with fewer than ten million employed. After categorizing the supersectors into two groups based on their COVID impact, the next step was to investigate which industries were the most affected by the pandemic. We hypothesized that the Leisure and Hospitality, and Transportation and Utilities industries would be hit hardest, while Private Education would be relatively unaffected. To validate this hypothesis, we conducted an Exploratory Data Analysis (EDA) using a line graph, which displayed the employment population of the affected supersectors for the years 2019 and 2020. The results of the EDA confirmed their hypothesis, with a significant decrease in employment population in the Leisure and Hospitality, and Transportation and Utilities industries from 2019 to 2020. Based on this confirmation, we decided to concentrate on analyzing the COVID affected group to gain further insights into the pandemic's effects. We believed that by focusing on the impacted supersectors and conducting statistical analysis, we would be able to validate our initial categorization of supersectors. This approach allowed us to narrow our research focus, enabling us to conduct a more detailed analysis of the industries that were most severely affected by the pandemic.

Statistical Analysis: ANOVA

The ANOVA table above shows the results of the statistical analysis conducted on the annual percent change as a function of being in the "Affected" group. It shows that there is a somewhat significant difference in the annual percent change between the "Affected" group and the "Unaffected" group. However, we were not satisfied with this level of significance.

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## affected     1   88.9   88.88    4.45 0.0477 *
## Residuals   20  399.5   19.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Statistical Analysis: ANOVA (contd.)

After adjusting the grouping, we created a new "Affected" group that included the Leisure and Hospitality, Retail Trade, Other Services supersectors, and the mining and logging subsector and moved the Transportation and utilities susperector to the "Unaffected" group.. This new grouping resulted in a much better fit with a highly significant p-value, providing support for our hypotheses. Specifically, we found support for the hypothesis that the Leisure and Hospitality supersector is in the affected group. On the other hand, we rejected the hypothesis that the Transportation and Utilities supersector is in the affected group. Finally, we found support for the hypothesis that the Private education and health services supersector is not in the affected group. These findings suggest that the new grouping of industries better captures the factors that are affecting changes in annual percent change.

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## affected     1  269.2  269.22   24.57 7.59e-05 ***
## Residuals   20  219.2   10.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary

Our study shows that the pandemic has had a severe impact on young workers aged 16-24, as they have faced higher unemployment rates compared to other age groups. This is likely due to the fact that younger workers are more likely to work in the hospitality and retail industries which were greatly affected by lockdowns and closures. These industries suffered significant losses and were forced to lay off many employees, leading to a sharp increase in unemployment rates in these sectors.

Additionally, we found that women had lower employment rates during the COVID-19 years compared to men, which may be due to the fact that women are overrepresented in the hardest hit industries. These findings highlight the need for targeted policies to support the most vulnerable groups, including young workers and women.

Finally, we found that minority groups, particularly Asian workers, experienced a disproportionate impact on employment rates during the pandemic. This is consistent with other studies that have shown that minority groups are more likely to work in industries that were most affected by the pandemic. These findings call for a closer examination of the impact of the pandemic on minority groups and the development of policies to support their recovery.

Future Studies:

For future work, we recommend further research into the effectiveness of policy interventions to mitigate the negative impact of the pandemic on employment rates. It would be useful to analyze the effectiveness of interventions and identify areas where improvements can be made. Another important area for future research is the impact of job loss and unemployment on mental health, including anxiety and depression. Previous research has shown that economic downturns can lead to negative mental health outcomes, and it would be interesting to see if this is also true for the COVID-19 pandemic. Finally, comparative studies on the impact of COVID-19 on the job market across countries and regions would also be valuable in understanding the global impact of the pandemic.

Sources

- Economic Policy Institute. (2020). The COVID-19 shock: A disproportionate and devastating effect on America's low-income workers. Retrieved from https://www.epi.org/publication/swa-2020-employment-report/
- Center for American Progress (2021)"How Covid-19 Sent Women's Workforce Progress Backward." Center for American Progress, 9 Nov. 2021, www.americanprogress.org/article/covid-19-sent-womens-workforce-progress-backward/.
- "Racial Disparities in COVID-19: Key Findings from Available Data and Analysis." KFF, 17 Aug. 2020, www.kff.org/racial-equity-and-health-policy/issue-brief/racial-disparities-covid-19-key-findings-available-data-analysis/.

-Honoré, B.E., Hu, L. The COVID-19 pandemic and Asian American employment. Empir Econ 64, 2053–2083 (2023). https://doi.org/10.1007/s00181-022-02306-5

- https://link.springer.com/article/10.1007/s00181-022-02306-5

Appendices

Appendix 1: Initial OLS Regression

```
Call:
lm(formula = Population ~ ., data = Demographics)

Residuals:
     Min       1Q   Median       3Q      Max
-3082.98  -471.21   -33.62   605.62  2361.66

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -232.7671   129.7731  -1.794 0.073213 .
raceBLACKAFRICANAMERICAN     319.6071    92.7345   3.446 0.000595 ***
raceHISPANICLATINOETHNICITY  605.5268    92.7345   6.530 1.11e-10 ***
raceWHITE                   3925.9821    92.7345  42.336  < 2e-16 ***
genderWOMEN                 -235.2723    65.5732  -3.588 0.000352 ***
Year2016                      28.9464   131.1464   0.221 0.825363
Year2017                      50.2679   131.1464   0.383 0.701593
Year2018                      79.2857   131.1464   0.605 0.545628
Year2019                     100.7679   131.1464   0.768 0.442479
Year2020                       0.6964   131.1464   0.005 0.995764
Year2021                      53.0000   131.1464   0.404 0.686216
Year2022                     118.1964   131.1464   0.901 0.367699
Age_Group25 to 34 years     1572.2995   111.9875  14.040  < 2e-16 ***
Age_Group35 to 44 years     1443.7135   111.9875  12.892  < 2e-16 ***
Age_Group45 to 64 years     1106.1276    93.6955  11.806  < 2e-16 ***
Age_Group65 to 74 years     -390.1146   111.9875  -3.484 0.000519 ***
Age_Group75 years and over  -684.4427   141.6543  -4.832 1.60e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 981.4 on 879 degrees of freedom
Multiple R-squared:  0.7697,    Adjusted R-squared:  0.7655
F-statistic: 183.6 on 16 and 879 DF,  p-value: < 2.2e-16
```

Appendix 2: t-test (gender)

```
        Welch Two Sample t-test

data:  c(mean_men_before, mean_men_after) and c(mean_women_before, mean_women_after)
t = 60.338, df = 1.8084, p-value = 0.0005347
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 216.8813 254.1187
sample estimates:
mean of x mean of y
 1676.977  1441.477
```

Appendix 3: Linear Regression and ANOVA (Age and Race)

```
Call:
lm(formula = Population ~ Age_Group + race, data = covid_data)

Residuals:
     Min       1Q   Median       3Q      Max
-2930.83  -476.17   -15.03   610.83  2179.95

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                   5053.0      206.7  24.443  < 2e-16 ***
Age_Group16 to 24 years      -1472.2      223.3  -6.593 3.28e-10 ***
Age_Group25 to 34 years        138.5      244.6   0.566   0.5718
Age_Group45 to 64 years       -354.0      211.8  -1.671   0.0961 .
Age_Group65 to 74 years      -1814.8      244.6  -7.419 2.69e-12 ***
Age_Group75 years and over   -2112.2      299.6  -7.051 2.39e-11 ***
raceASIAN                    -3895.8      184.9 -21.070  < 2e-16 ***
raceBLACKAFRICANAMERICAN     -3580.9      184.9 -19.367  < 2e-16 ***
raceHISPANICLATINOETHNICITY  -3286.1      184.9 -17.772  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 978.4 on 215 degrees of freedom
Multiple R-squared:  0.7714,    Adjusted R-squared:  0.7629
F-statistic:  90.7 on 8 and 215 DF,  p-value: < 2.2e-16

Analysis of Variance Table

Response: Population
           Df    Sum Sq   Mean Sq F value    Pr(>F)
Age_Group   5 143578632  28715726  29.997 < 2.2e-16 ***
race        3 550989241 183663080 191.861 < 2.2e-16 ***
Residuals 215 205813650    957273
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```