

**Census API Data - Annual Business Survey (ABS) 2019**

**Salma Tahlil, Rob Dewan, Susan Lu, Mason Lonoff**

**Dev 10**

## Data Sources:

Bureau, U. S. C. (2022, October 28). *Annual Business Survey (ABS) apis*. Census.gov. Retrieved December 29, 2022, from <https://www.census.gov/data/developers/data-sets/abs.2019.html>

*Racial underrepresentation in construction: How do the union and nonunion sectors compare?* Economic Policy Institute. (n.d.). Retrieved January 2, 2023, from <https://www.epi.org/blog/racial-underrepresentation-construction/>

*When It Comes To Automation, Not All Industries Are Equal*. Forbes. (2020, Dec 4). Retrieved December 28, 2022, from [When It Comes To Automation, Not All Industries Are Equal \(forbes.com\)](https://www.forbes.com/sites/forbesrealestatecouncil/2020/12/04/when-it-comes-to-automation-not-all-industries-are-equal/)

## Initial Questions:

- How have industries changed as automation has increased?
- What industries have a higher number of black owners? Are they different or similar to female black owners? Are there drastic differences between the two? How different are they from white owners? (all in the year 2020)
- What educational background do the business owners in America have and do they vary significantly by race? What motivates them to own their own business?

### ***Annual Business Survey (ABS)***

The Annual Business Survey conveys information on economic and demographic characteristics for businesses and their owners by sex, ethnicity, race, and veteran status in the United States. The survey also provides information on company summary and module business characteristics. The Annual Business Survey data was collected from online based surveys and data from Economic Census as well as administrative records. A random sample of businesses were selected and sent letters which gave instructions on how to report online.

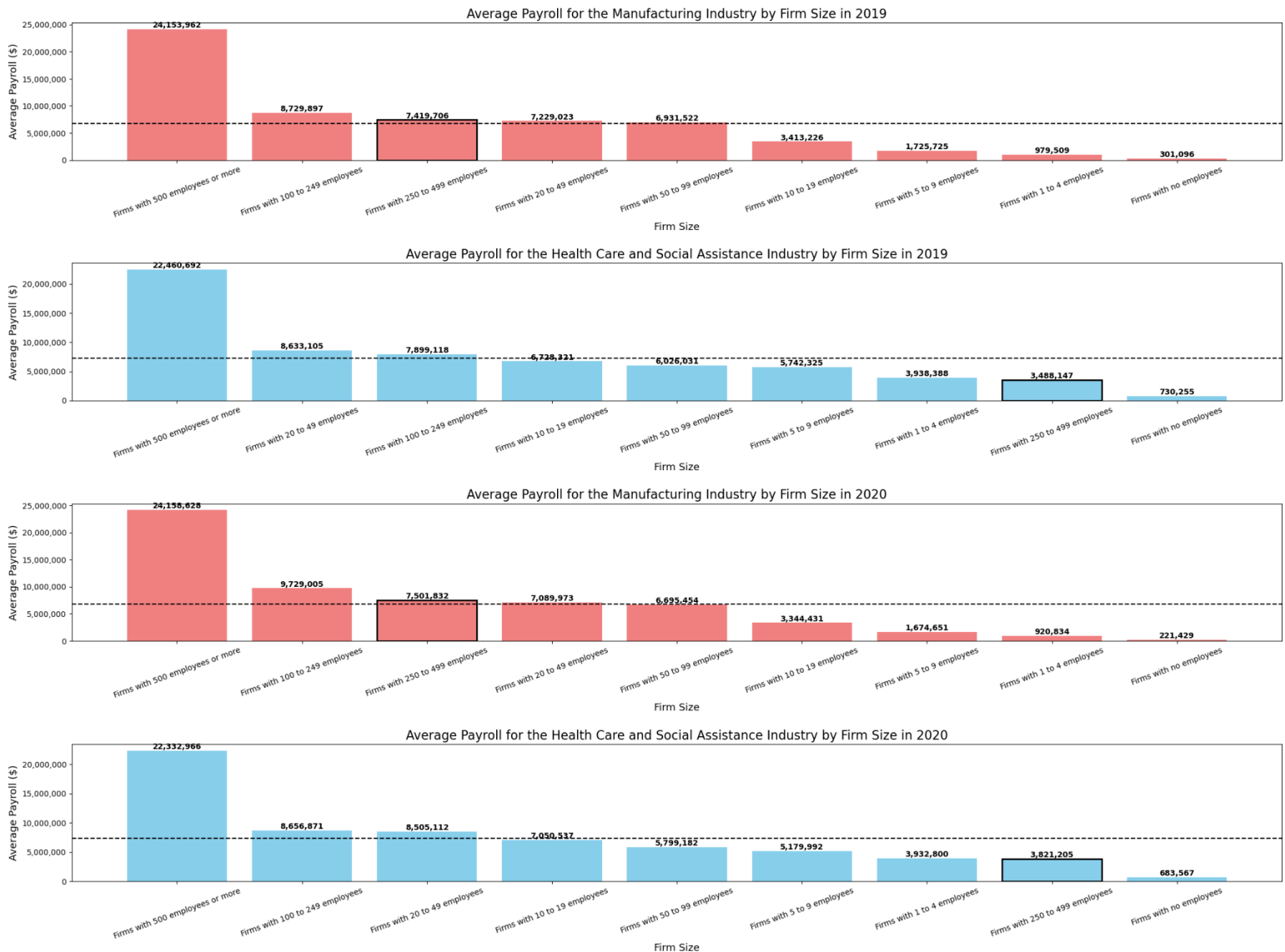
### ***Company Summary***

*Company Summary* is also a dataset provided by the Annual Business Survey (ABS). This dataset contains information concerning general information about businesses. For example, it has information about sales, payroll, number of employees, and years in business. Similar to the other datasets, *Company Summary* also contains categorical data about business owners such as their race, sex, ethnicity, and veteran status. While there are many years to analyze, we are only concerning ourselves with the data from the years 2019 and 2020.

Prior to any type of analysis being performed, it was important to brainstorm certain interesting angles to consider for further analysis. Our team was aware of the recent and increasing trend of automation in business. We decided that examining how different levels of automation have affected businesses could yield interesting results. The first step was to figure out which industries were on each end of the automation spectrum. Well into the 21st century, it is widely known that the Manufacturing Industry has adopted many technological advancements to increase automation. However, certain industries have not been automating at the same pace. According to an article by Forbes, only 20% of healthcare executives claim to have invested in major automation improvements (ServiceNow). Our team felt confident that comparing these two industries could yield some interesting results.

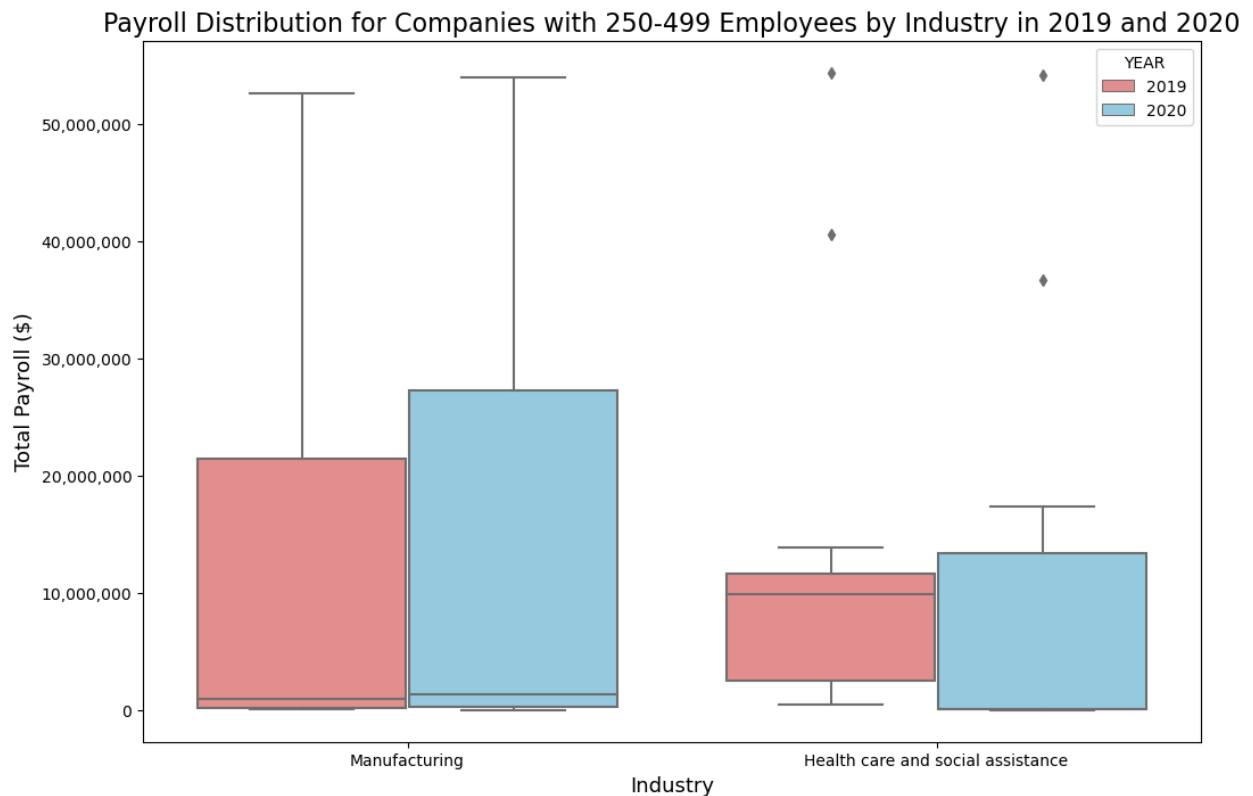
Our team conducted exploratory research on these two industries. We created a correlation matrix and found that payroll and employee size had an extremely strong positive correlation at 0.980673. To double check that our data was showing the same correlation, we visually compared Average Payroll by Number of Employees. It is important to note that the x-axis is not the same column of data that we ran the correlation on. The difference between the two columns is that our x-axis groups the employee sizes into bins whereas the data used in the correlation matrix is a raw number. Below is our first visualization. Based on our prior exploratory research, we would expect to see the highest average payrolls be the largest companies. For the Manufacturing Industry, that logic mostly tracks; while it's not exactly largest to smallest, the three largest company sizes have the three largest payrolls. However, in the Health Care and Social Assistance Industry, our 2nd largest company size (250-499 employees)

has the 2nd lowest average payroll for both 2019 and 2020. Our team decided that further analysis on this was necessary.

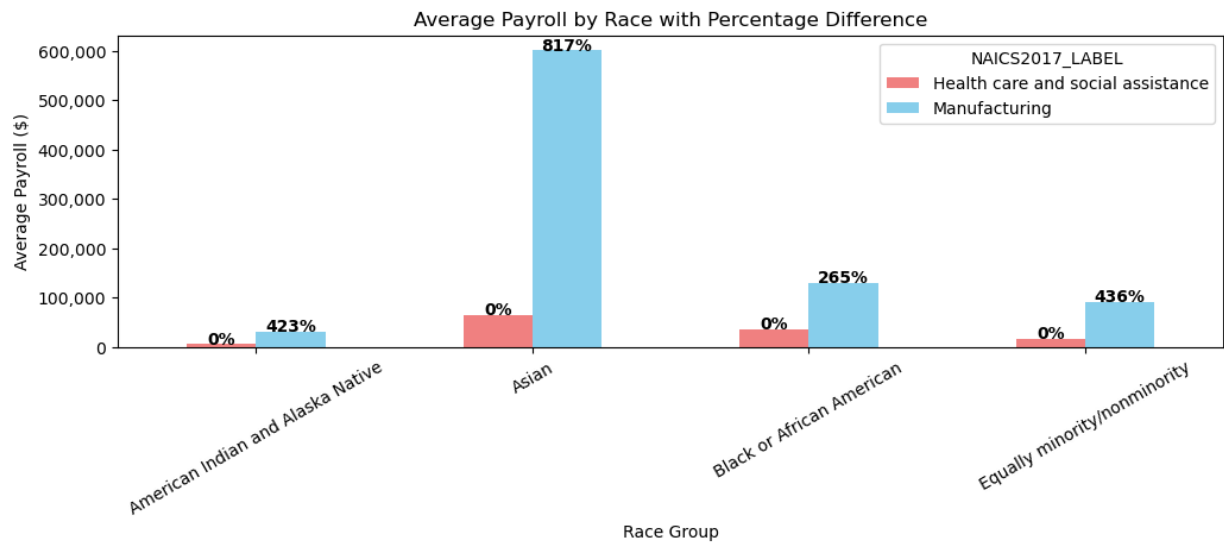
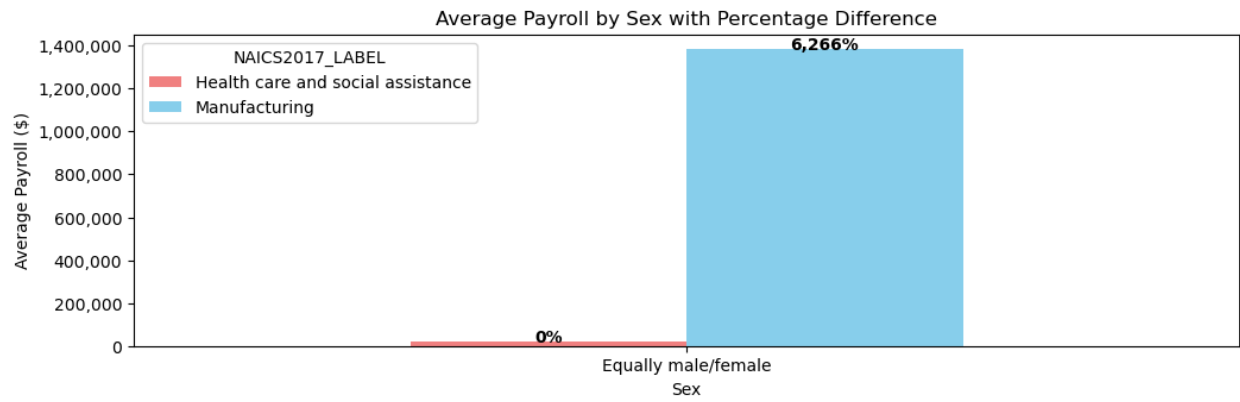


The next step in our analysis was to create a box and whisker plot of the payrolls. One of the first aspects that stands out are the outliers seen with the Health Care and Social Assistance Industry. Those outliers strongly affected some of our descriptive statistics as, on average, the skewness score for the Health Care and Social Assistance Industry is roughly double (2.395 vs 1.22) the skewness score for the Manufacturing Industry's payroll data. Besides the outliers, it is clear that our upper quartile is considerably lower in the Health Care and Social Assistance Industry. Also, the standard deviation for the Manufacturing Industry is \$3,654,111.5 more than in the Health Care and Social Assistance Industry. We could then conclude that the data is more

tightly grouped in the Health Care and Social Assistance Industry. While important insights were found, no conclusions as to why the payrolls differ could be formulated.



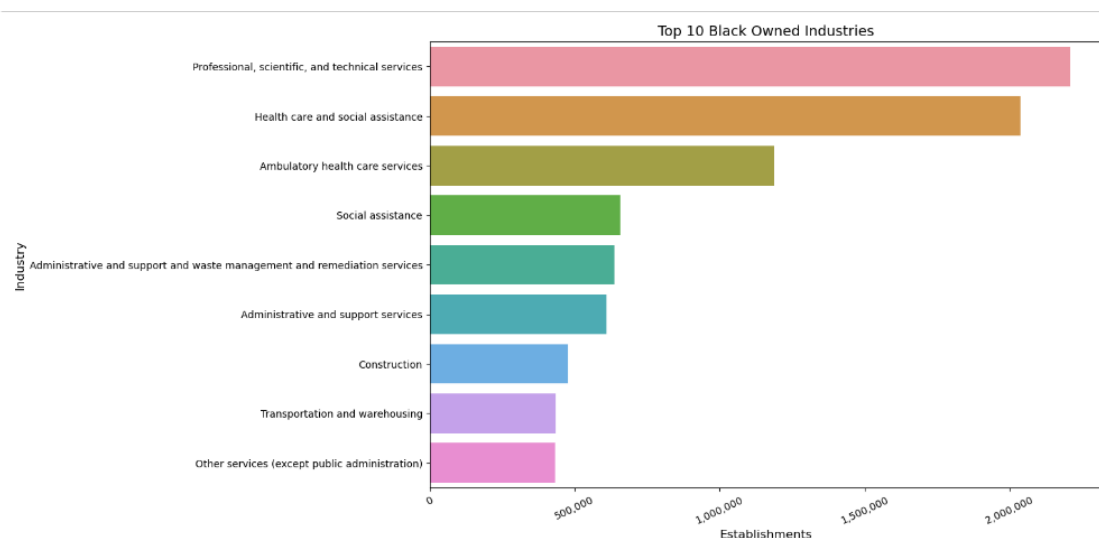
Our team decided that the next step was to examine the categorical differences between the business owners. Specifically, we wanted to narrow down exactly where the pay difference was the most pronounced. On average, for the companies with 250-499 employees, companies in the Manufacturing Industry paid their employees 113.46% more than the companies in the Health Care and Social Assistance Industry. Our team then went through each category and calculated the percent difference in pay between the two industries. Our final visualization contains the top five largest percent differences in pay between the industries based on some categorization (race, sex, ethnicity, and/or veteran status). Unfortunately, no concrete conclusions can be made using this dataset as the data is too general. These five owner categories represent exactly where our team would want to dive into deeper analysis.



### *Characteristics of Businesses*

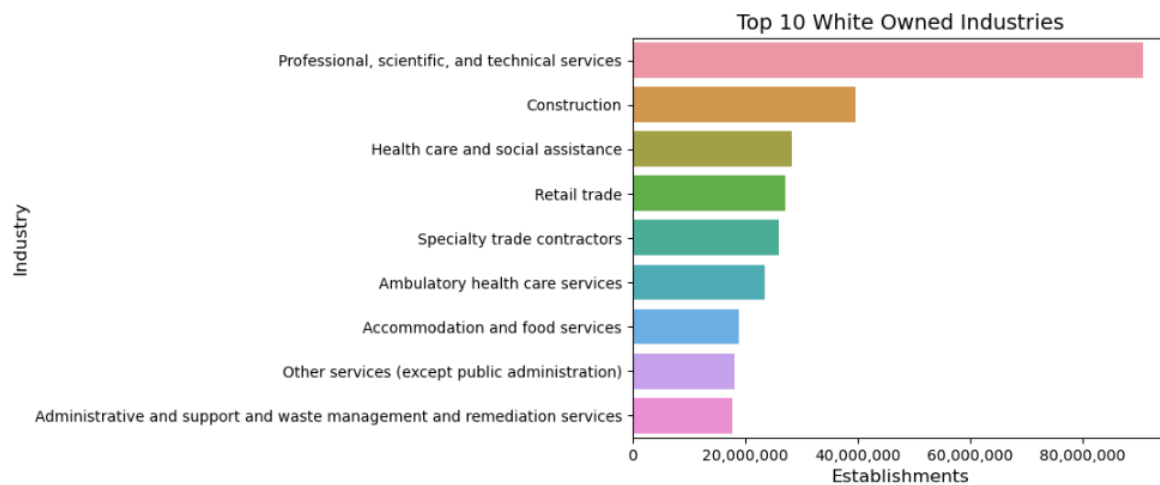
*Characteristics of Businesses* is another dataset provided by the Annual Business Survey (ABS). This dataset provides data for the respondent employer firms by categories such as sector, sex, ethnicity, race, veteran status, years in business etc for the United States. Some of the years go further in depth providing specific states and metro areas such as 2018 and 2021. The year being analyzed for this dataset is 2020 which includes survey responses from the previous year. After exploring the data, questions arose such as what industries have a higher number of black owners? How different are they from white owners? Are there drastic differences between the two? Are they different or similar to female black owners?

The main reason to further explore race and the type of industries owned was to see if there were any significant differences. We hypothesized that black owned and white owned industries would be drastically different due to trends that have been observed over history. Black people have experienced years of discrimination and unequal opportunities compared to their white counterparts. When analyzing the first graph, there is a large amount of black owners in professional, scientific, and technical services and healthcare and social assistance.



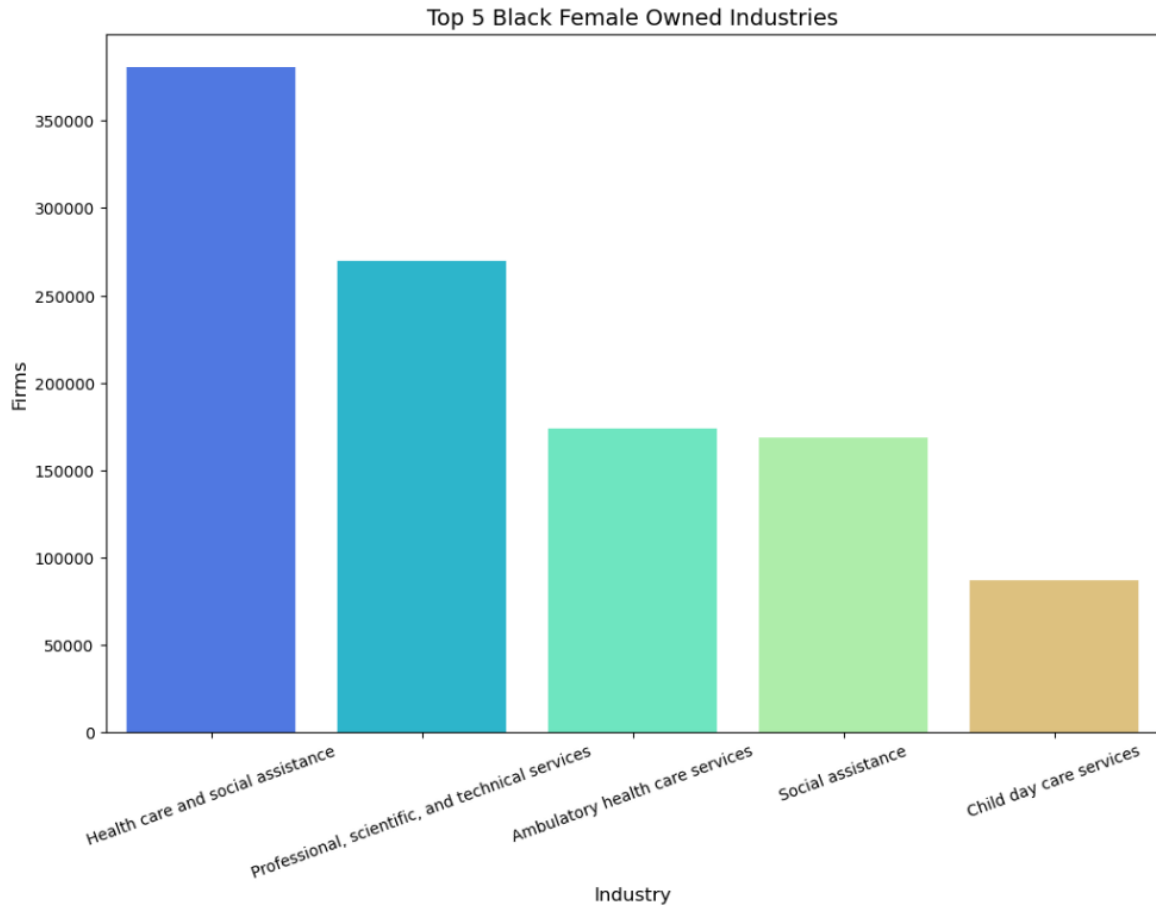
To compare the first graph to the second, we looked at white owned industries. What was really surprising was that both black and white owned industries' highest number of establishments were professional, scientific, and technical services. However, there was a significant disparity in the number of industries held. Industries owned by White Americans outnumbered those held by black people by a large margin. Despite Black Americans exceeding in industries in the technical fields, there is still a discrepancy between the different races. Another major difference between the two graphs is how the second largest industry category was healthcare in the black owned graph and was construction for the white owned graph. After some further investigation, we discovered that construction is a sector of the American economy that has historically excluded black workers(*Racial underrepresentation in construction: How do*

*the union and nonunion sectors compare?* Economic Policy Institute, n.d.). Some of the reasons behind this discrimination are said to be dated back to Jim Crow laws (*Racial underrepresentation in construction: How do the union and nonunion sectors compare?* Economic Policy Institute, n.d.). These findings gave reason behind the numbers indicated in the graph and explains why black owners in the construction industry are notably lower. Another notable difference in the graphs is black owners are more in the health sector whereas white owners are technical, construction and retail.



The last graph observed female black owners. Before making the graphs, we hypothesized that female owned industries would be significantly less than the other two graphs. This is also due to the other graphs not being broken up by gender. We wanted to look into black female owned industries and inferred it would be mostly healthcare. After making the graphs it shows the highest amount of ownership in healthcare and social assistance. This was influential to observe because black people need to experience seeing people in healthcare that look like them. This could help bridge the gap and the racial divide in healthcare. It also shows child care which was not a top industry in the other two graphs. We were surprised to see the second industry in professional, scientific, and technical services.



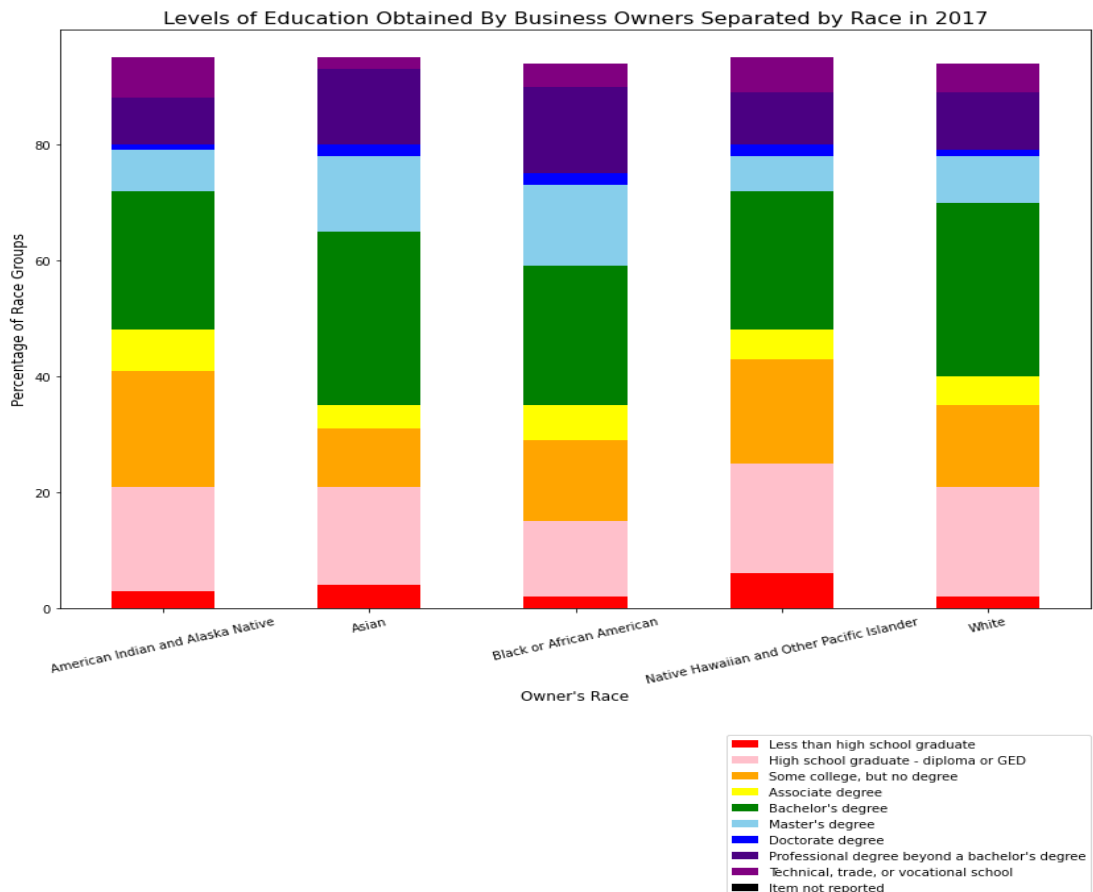


To conclude, after further consideration this dataset provided useful information but there were limitations. The dataset we worked on was already aggregated so it was difficult coming up with accurate conclusions. Although there were limitations, these three graphs provided insight on race and areas in which there are disparities within different industries. Even though the United States has shown efforts in improving race disparities there are still industries that need to improve and be more racially inclusive.

## Characteristics of Business Owners

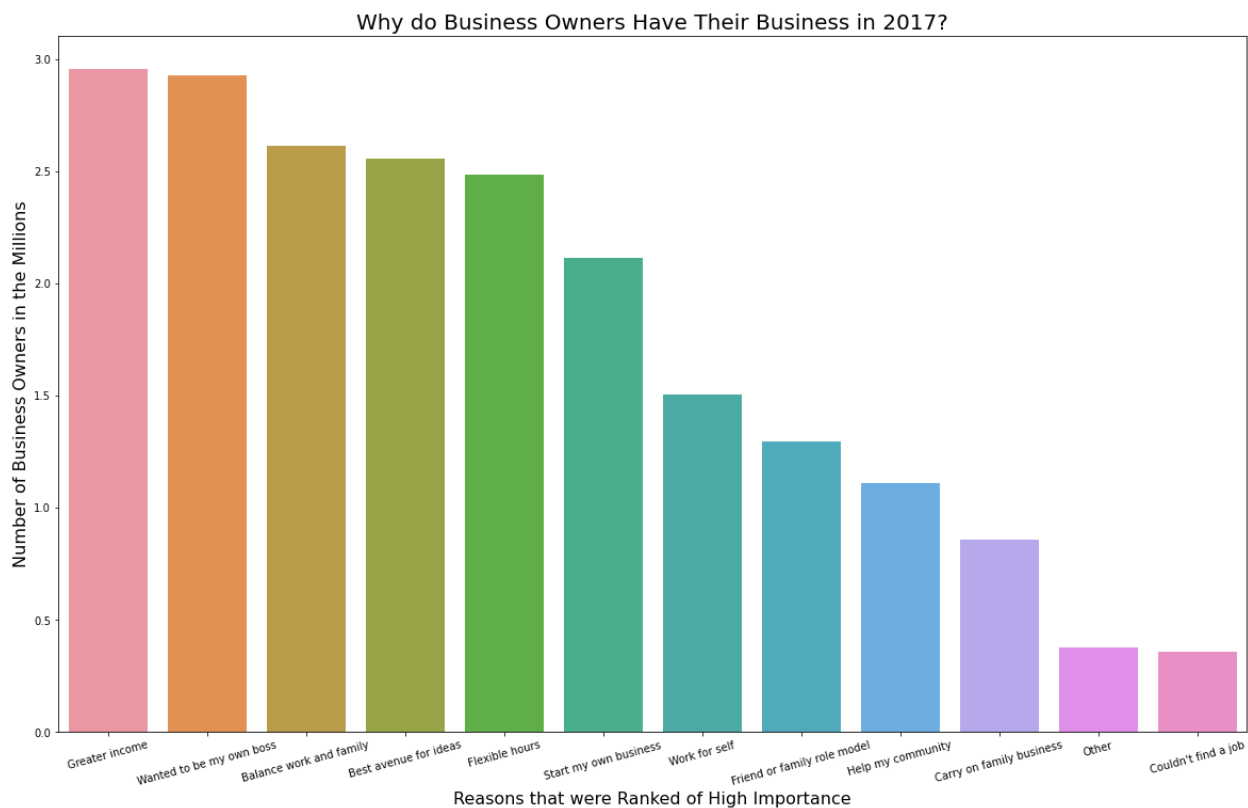
To provide a deeper look into our business owners, the *Characteristics of Business Owners* dataset gathers information specifically pertaining to the business owners, such as level of education completed, reasons for the acquisition of the business and how the business was acquired. The dataset for 2017 was analyzed because there were more questions on that specific survey compared to the following three years. It is worth noting that the data provided is an aggregated set of data, which limits our ability to correlate two variables. However, we are still able to make observations about our business owners as a whole.

So who are our business owners? According to this dataset, there were approximately 4.74 million business owners, predominantly of white race. It was hypothesized that the highest level of education completed would vary across different races. Due to the population of white business owners being significantly higher than all other recorded races, the completed education levels were graphed as a percentage. The number of business owners were aggregated together by business sector and race and then divided by the total number of business owners for each race category. Unfortunately, the percentages were rounded down so the total for each category came out to about 94 - 95%. But even not having a total of 100%, this still allowed for easy comparison of how each race varied in education.

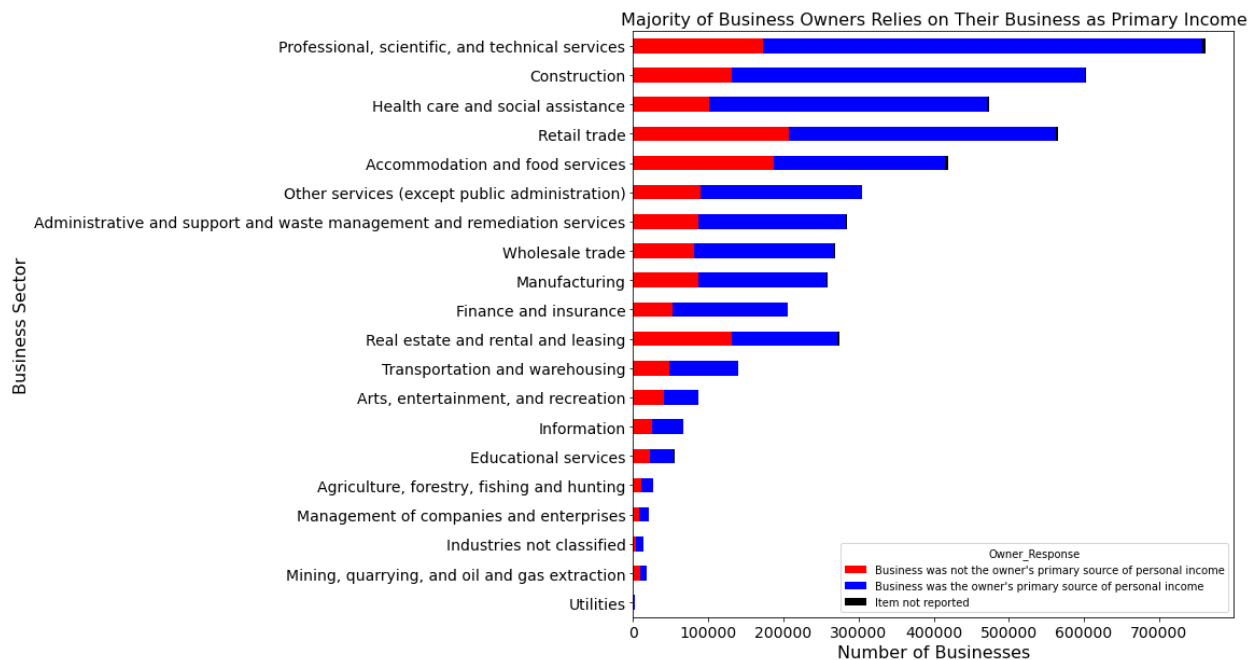


The bar graph above shows across every race, at least 30% of each race category did not have a college degree. Technical, trade or vocational school seems more popular among the American Indian and Alaska Natives and Native Hawaiian and other Pacific Islander races. The Master's degree, on the other hand, appears to be more popular among the Asian and Black or African American races. We also see that the percentage of business owners who received an Associate degree is much lower than those who received a Bachelor's degree. This might mean that the hurdle of completing a Bachelor's degree compared to getting an Associate's degree is not significant. However, there is quite a decrease when comparing the completion of a Bachelor's degree compared to the completion of a Master's degree. However, this might be due to business owners deciding to pursue professional degrees beyond a Bachelor's.

It is a common saying that owning a business is hard. So why did our business owners take on the challenge? What values drive them to run a business? The graph below shows the reasons that business owners have ranked very important to them as to why they run a business. The most common reason appears to be for a greater income, with wanting to be their own boss as a close second. The least common reason is the inability to find a job. It was very interesting that about 400 thousand business owners were unable to find jobs and therefore, created their own.



Given that a greater income was the most common reason why a person would own a business, it was hypothesized that perhaps the business income was treated as an additional income, and not the primary income. However, based on the following graph, more than half of the businesses in just about every industry are the sources of the owner's primary income. This does support the second most common motivation for business owners of wanting to be their own boss. We also notice that some industries, such as Arts, Entertainment and Recreation have more of an even split between those who rely on the business for primary income, while some of the other industries like Professional, scientific and technical services have over three times as many owners who rely on the business for primary income compared to those who do not.



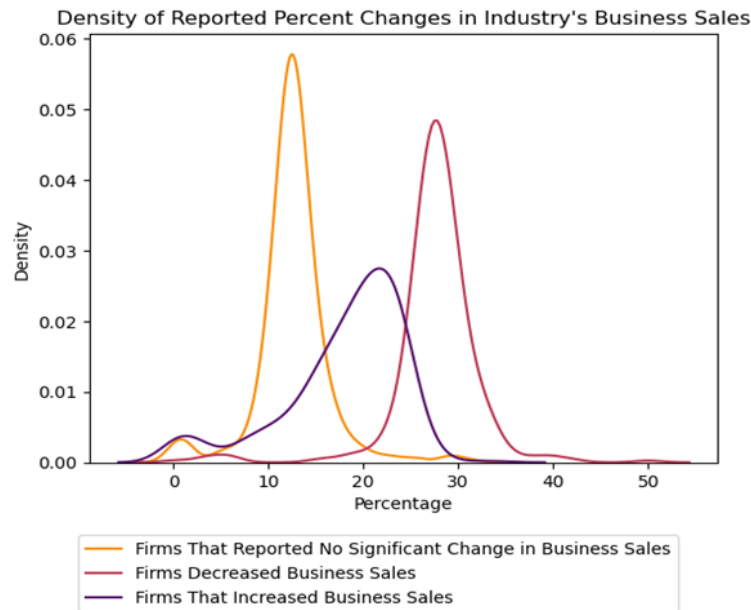
Based on the 2017 dataset for Characteristics of Business Owners, we have observed that having a college degree might not be necessary to become a business owner. If we combine all the categories without a college degree, it would cover the largest percentage across every race. However, this is likely to be different depending on the business industry. If one wants to determine what level of education is common for a particular industry, one can isolate data for that particular industry.

We can also conclude that most business owners find having their own business is financially advantageous since many of them became business owners because they want to achieve greater income. We have also seen that the majority of the business owners rely on this greater income as their primary income. It also suggests that working for someone else would result in a lower income and is not as satisfying as being your own boss. Even though owning a business is hard work, perhaps the reward makes it worthwhile.

## ***Module of Business Characteristics***

### ***Distribution of Industry Business Sales***

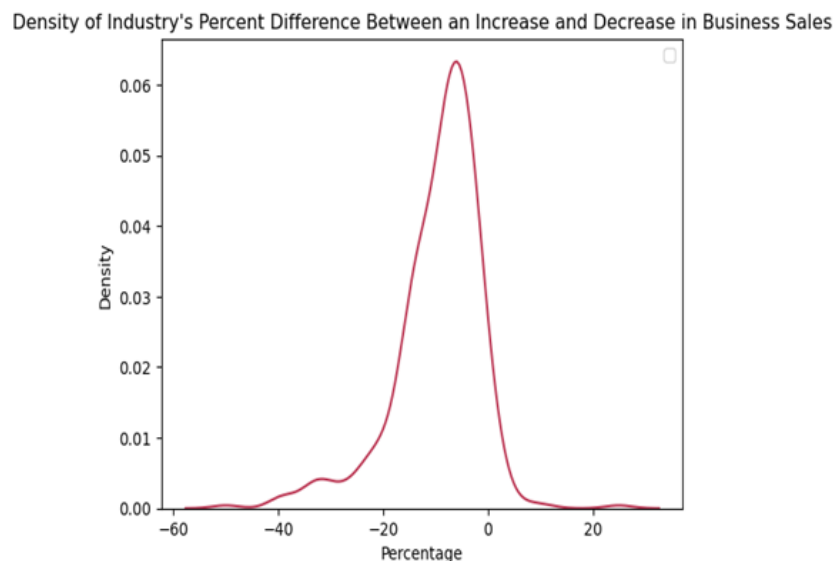
The *Module of Business Characteristics (MBC)* is a dataset provided by the Annual



Business Survey (ABS) which is accessible via an API provided by the [US Census Bureau](#). The data set, conducted in 2021, covers a wide range of survey topics on the business' use of technology and the financial effects of the COVID-19 pandemic. In this section of the report, we will be focusing on how the pandemic affected the business sales of various industries.

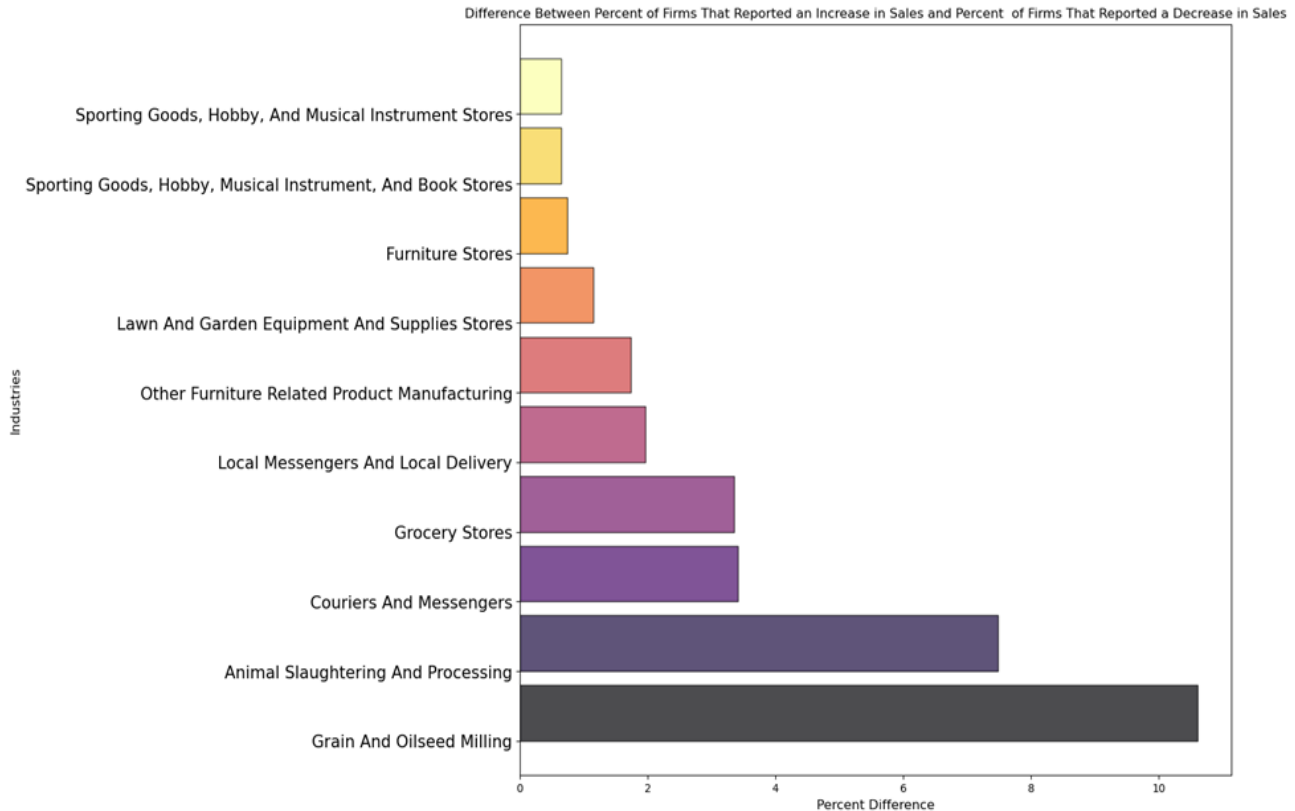
The *MCB* assessed the financial wellbeing of firms through the following question: "How would you assess the overall effect of the coronavirus pandemic

on this business's sales in 2020?" Intuitively, anyone who endured the pandemic may be predisposed to believe that statewide shutdowns uniformly impacted business sales for the worse; however, for a handful of industries, this was not the case. Plotted immediately left are the density distribution curves for the percentage of firms in each industry that either reported that their sales increased, decreased, or experienced no significant change. Meanwhile, plotted immediately beneath the previous graph, we see that the percent difference between firms that reported an increase in business sales and firms that reported a decrease in business sales is seldom greater than zero. This implies that, among the sample of industries which responded to the *MCB*, most industries experienced a decline in business sales.

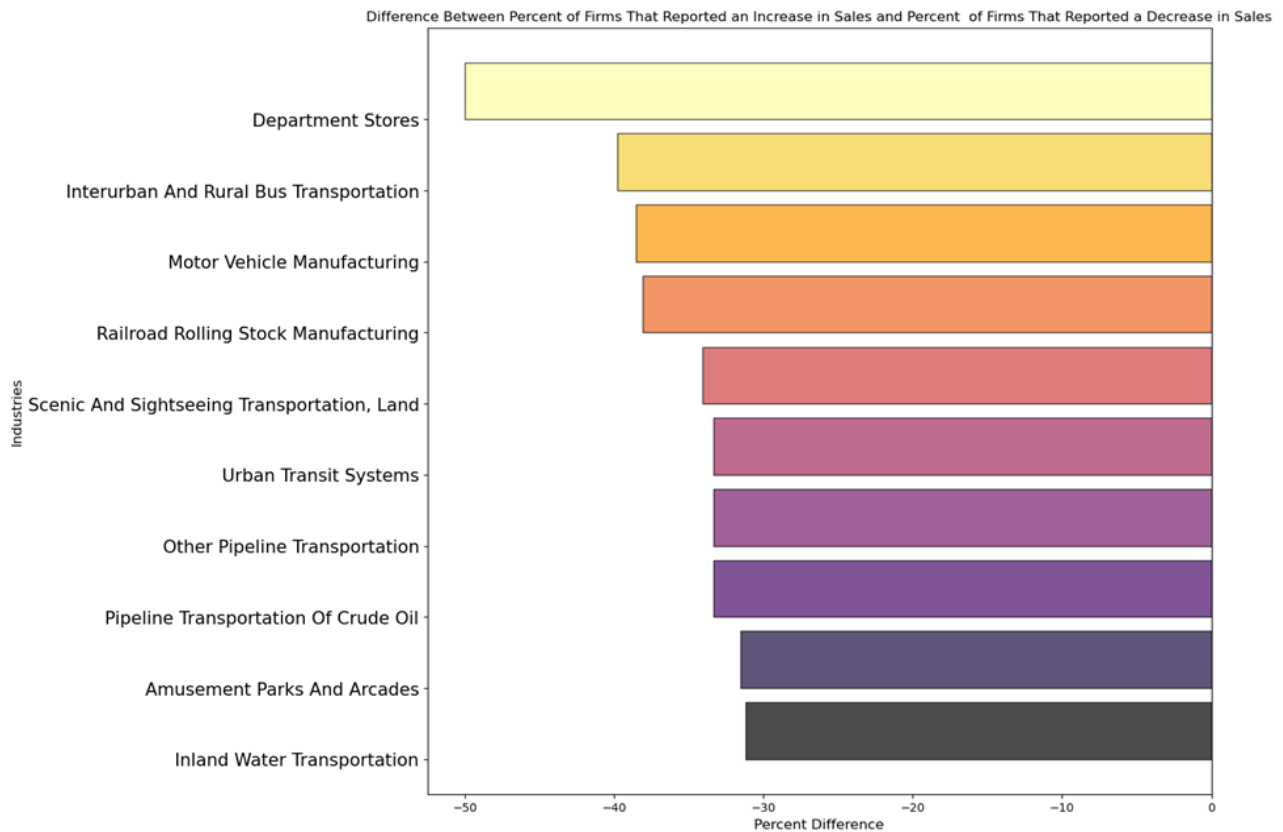


### The Chosen Few

Although, as alluded to earlier, this trend of declining business sales was bucked by a smattering of firms in some industries. Plotted below are the top ten industries where the percentage of firms that reported an increase in business sales exceeded the percentage of firms that reported a decline in business sales. By the *MCB*, while most businesses experienced a decline in business sales, the pandemic was exceptionally kind to industries who's primary concerns were either animal slaughtering or processing and grain and oilseed milling.



Meanwhile, by the same metric, we can identify the industries for which firms experienced a decline in sales precipitously greater than the percentage of firms that reported an increase in sales. From the gathered data we observe that department stores were, by the *MCB*, an industry most likely to experience a decline in sales.



## Conclusion

Parsing through the *MCB* dataset we have observed that most firms believe that the COVID-19 pandemic led to a decline in business sales. However, we also observed that this trend did not hold for a subset of industries.

However, there is an important discussion to be had in recognizing the limitations of this study. While the *MCB* covers a wide range of topics, only one question in the survey was explored here. To capture a clearer picture of the financial effects of the COVID-19 pandemic on specific industries, this study can be further enhanced by digesting the complete dataset.

Further, the methodology by which the percentage of the number firms for each industry is determined relies upon the assumption that this number can be found by totaling the number of observations corresponding to a question asked by the ABS. Should this assumption be proven false, much of this study may need to be refactored.

In conclusion, though most firms reported a decline in sales, some were fortunate enough to have their sales stagnate or increase due to the COVID-19 pandemic. Perhaps in researching this pandemic's effect on business we can further prepare ourselves for the next pandemic.