

Final Report

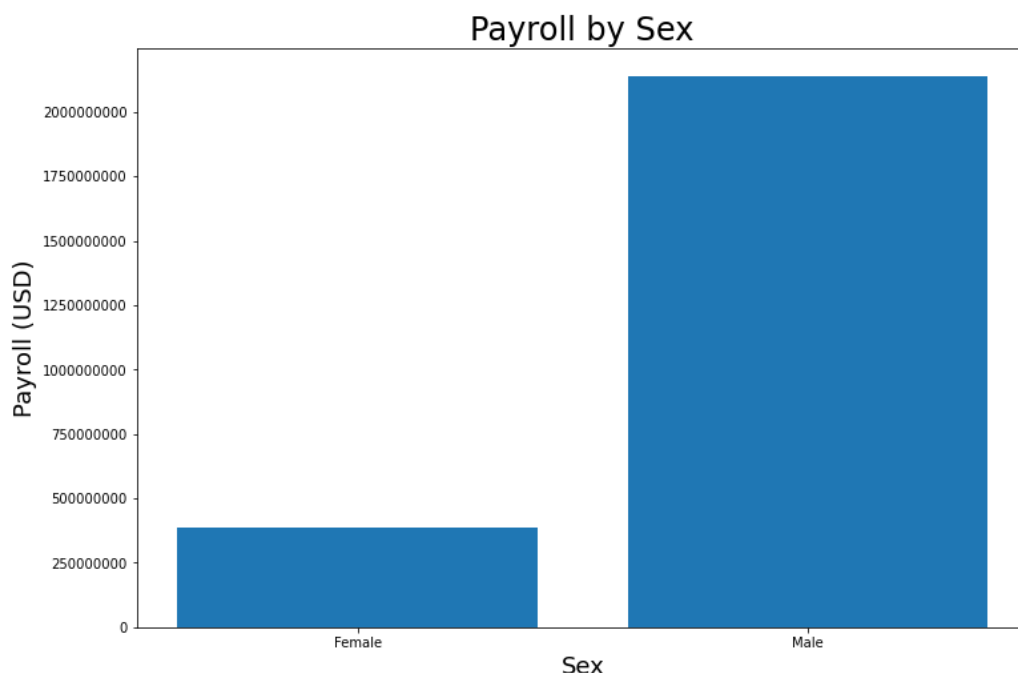
Dean Papadopoulos, Philip Waymeyer, Juliann Groglio, Savion Ponce

7/18/22

For this analysis, the data used was from the [Annual Business Survey 2019](#) from the United States Census Bureau. This survey includes information about the financials and demographics of business owners in the United States. The datasets used are Company Summary and Technology Characteristics of Businesses. The Company Summary dataset focuses on demographics of business owners such as sex, ethnicity, race, and veteran status. It also includes information on financials and other business information such as annual payroll, revenue, firm size, and number of employees. The Technology Characteristics of Businesses dataset focuses on the same demographic information along with technological information such as technology use, production, affects, and motivations. After reviewing this data, questions were raised on the difference in payrolls and revenues between demographics and the correlations between firm sizes and revenues. The following graphs and their corresponding descriptions display the findings from this data and the conclusions that were able to be confidently made.

The first three graphs represent an analysis of annual payroll versus different demographic categories such as sex, ethnicity, and race. For clarity, annual payroll is the amount of money a business owes its employees that includes payables such as wages, bonuses, and benefits. The question leading to the first graph asks, “How do the

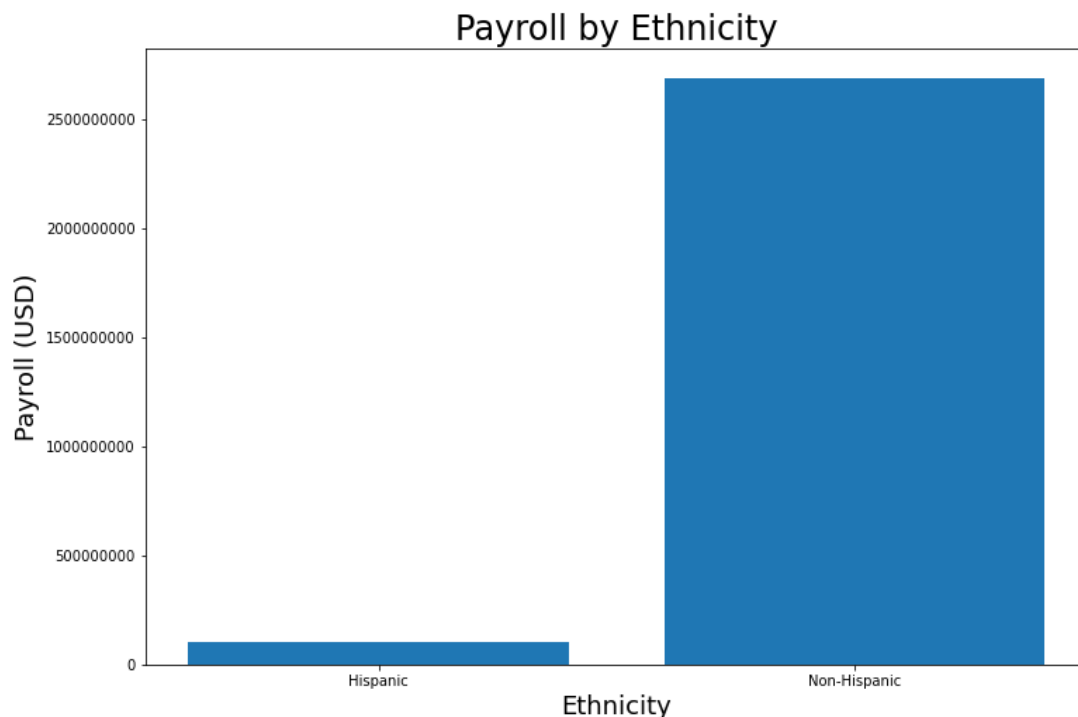
payrolls of males and females of all demographics differ?” This question required creating variables that only stored the total males and total females with all other demographics such as ethnicity and race set to their totals as well. This exposed the annual payrolls of males and females of no specific other demographic, making it simple to compare and analyze them. These variables along with the ‘Payroll’ column were used to create a bar chart using the Matplotlib library to plot the data. The graph created follows:



This graph shows the two sexes, male and female, on the x-axis, and the corresponding annual payrolls on the y-axis in USD, ranging from \$250,000,000 (\$250M) to \$2,000,000,000 (\$2B). The female business owners come in at around \$388,000,000 (\$388M), and the male business owners come in at around \$2,100,000,000 (\$2.1B). So, going back to the originally posed question of ‘How do the payrolls of males and females of all demographics differ’, the answer found is that male

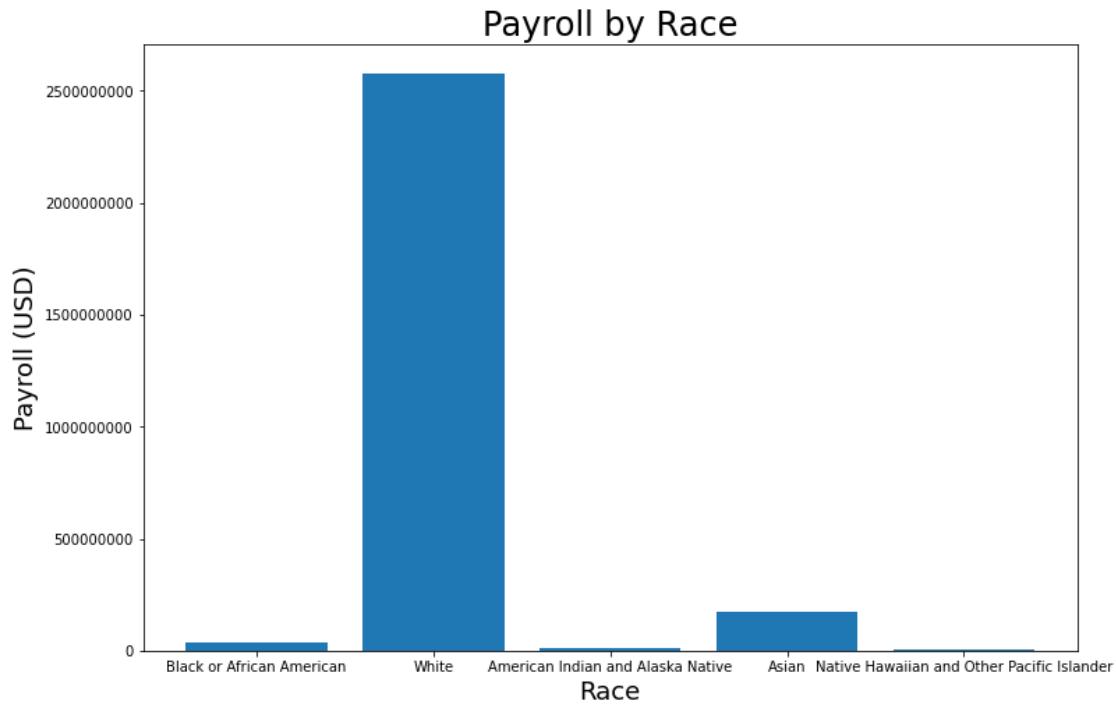
business owners have an annual payroll of more than four times the female business owners' annual payrolls.

The question leading to the second graph asks, “How do the payrolls of different ethnicities of all other demographics differ?” This question required creating variables that only stored the total non-Hispanic and total Hispanic business owners with all other demographics such as sex and race set to their totals as well. This exposed the annual payrolls of non-Hispanics and Hispanics of no specific other demographic, making it simple to compare and analyze them. These variables along with the ‘Payroll’ column were used to create a bar chart using the Matplotlib library to plot the data. The graph created follows:



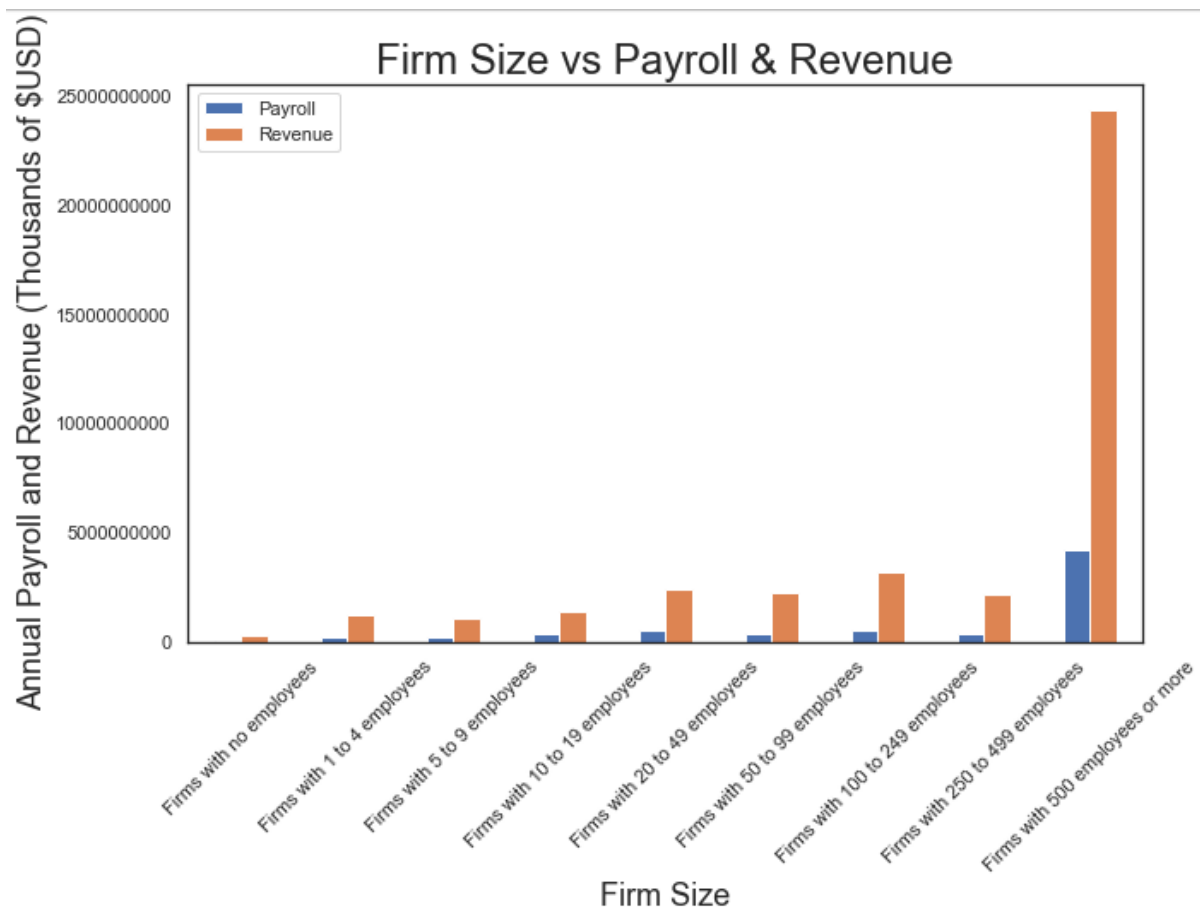
This graph shows the two ethnicities, non-Hispanic and Hispanic, on the x-axis, and the corresponding annual payrolls on the y-axis in USD, ranging from \$0 to \$2,500,000,000 (\$2.5B). The Hispanic business owners come in at around \$100,000,000 (\$100M), and the non-Hispanic business owners come in at around \$2,600,000,000 (\$2.6B). So, going back to the originally posed question of 'How do the payrolls of different ethnicities of all other demographics differ', the answer found is that non-Hispanic business owners have an annual payroll of more than twenty times the Hispanic business owners' annual payrolls.

The question leading to the third graph asks, "How do the payrolls of different races of all other demographics differ?" This question required creating variables that only stored the total Black/African American, White, American Indian/Alaskan Native, Asian, and Native Hawaiian/other Pacific Islanders business owners with all other demographics such as sex and ethnicity set to their totals as well. This exposed the annual payrolls of all five different races of no specific other demographic, making it simple to compare and analyze them. These variables along with the 'Payroll' column were used to create a bar chart using the Matplotlib library to plot the data. The graph created follows:



This graph shows the five races, Black/African American, White, American Indian/Alaskan Native, Asian, and Native Hawaiian/other Pacific Islanders on the x-axis, and the corresponding annual payrolls on the y-axis in USD, ranging from \$0 to \$2,500,000,000 (\$2.5B). The Black/African American business owners come in at around \$36,000,000 (\$36M), the White business owners come in at around \$2,500,000,000 (\$2.5B), the American Indian/Alaskan Native business owners come in at around \$8,000,000 (\$8M), the Asian business owners come in at around \$175,000,000 (\$175M), and the Native Hawaiian/other Pacific Islanders come in at around \$2,000,000 (\$2M). So, going back to the originally posed question of 'How do the payrolls of different races of all other demographics differ', the answer found is that White business owners have an annual payroll of more than ten times all the other business owners' annual payrolls combined.

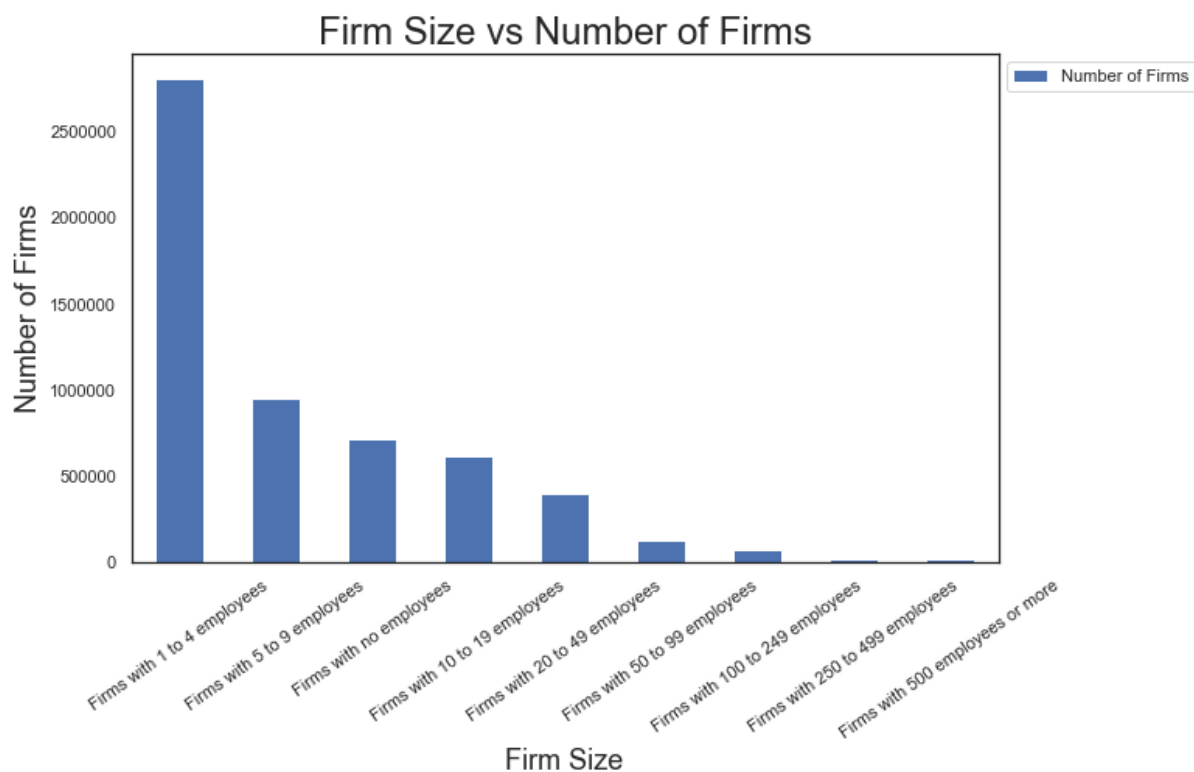
The question leading up to the fourth visualization asks, “Does payroll and revenue consistently increase with firm size?” In order to create a respective visualization, a data frame which included total sex, ethnicity, race, and veteran status was required. After this all variables except for firm size, payroll, and revenue should be filtered out. This final data frame was used to create the following visualization via matplotlib.



The visual above can help one understand if payroll and revenue increase as firm size increases. The findings from this graph would be vital to any business owner looking to expand their firm size. In this bar graph firm size is broken down into 9 different categories ranging from firms with zero employees to firms with 500 or more

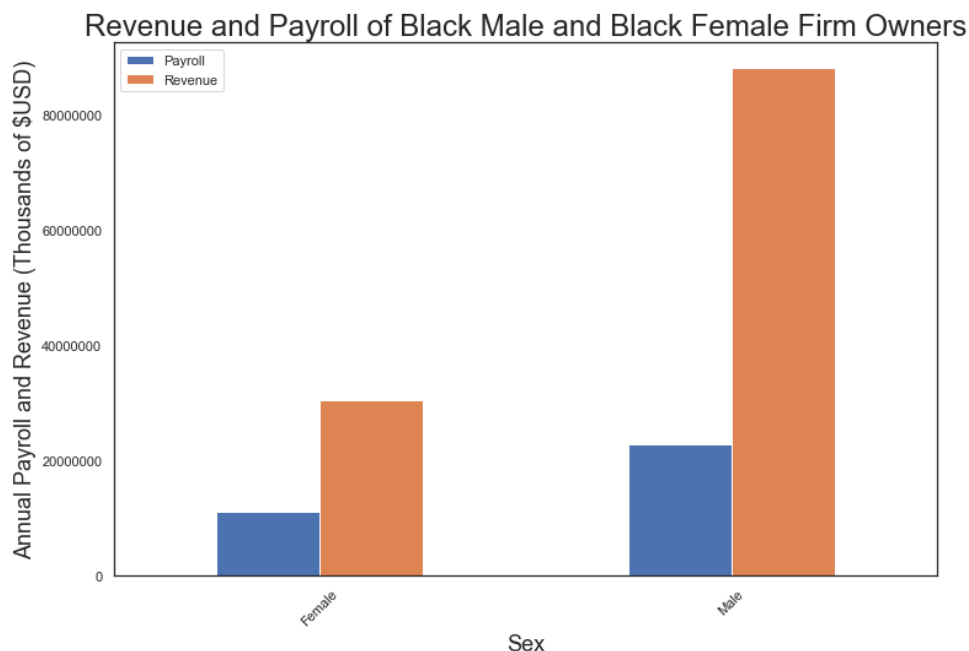
employees. The y axis represents dollars ranging from 5 trillion to 250 trillion. From this graph we can conclude that there is not a consistent increase in revenue and payroll as firm sizes increase. This is seen in the difference between the firm sizes of 100-249 employees and 250-499 employees. Surprisingly, this increase in firm size represents a drop in both payroll and revenue. Based on this visualization depending on a firm's size they may be cautious or encouraged to expand employment.

The question leading to the fifth visualization asks, “How many firms are there of each size?” Creating this visualization required similar steps to that of the visualization above. First a data frame where sex, ethnicity, race, and veteran status were set to total was required. After this the only variables needed in this data frame were firm size and number of firms. Once this final data frame was made it was used to make a bar graph via seaborn.



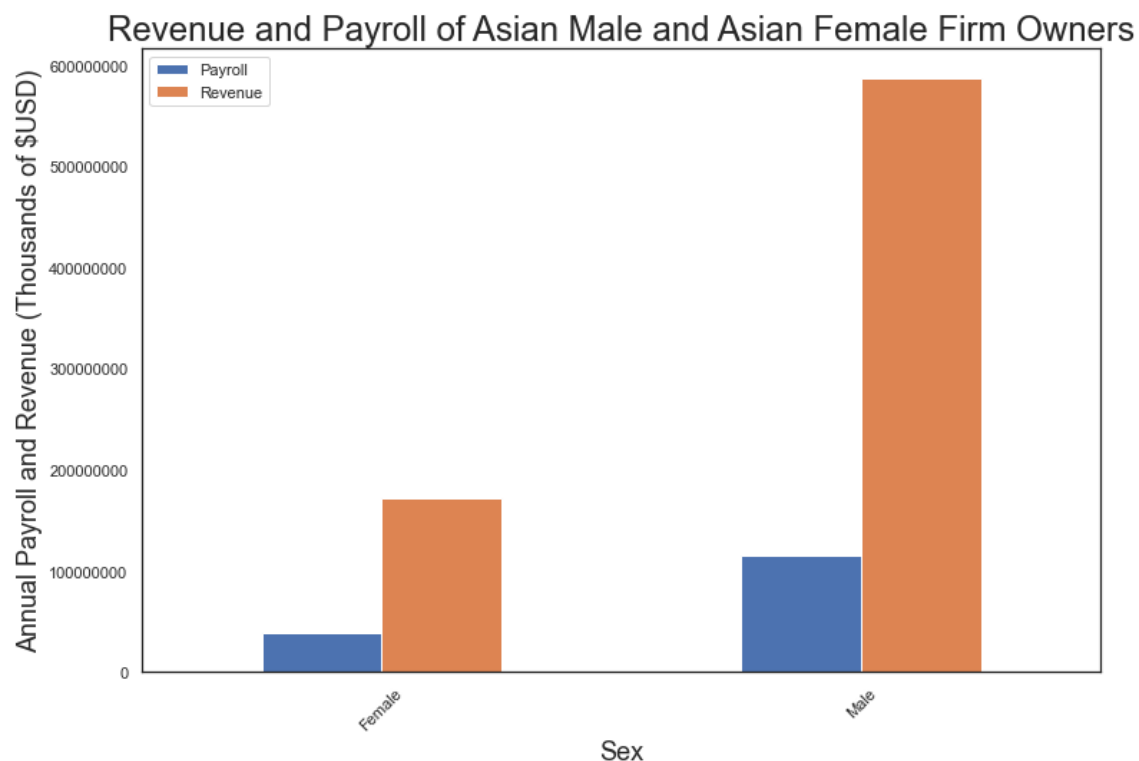
The graph above displays the categorical values of firm size on the x axis and the number of firms on the y axis. From this visualization it can be concluded that the most common firm size is 1-4 employees while least common is 500 or more employees. This is telling of how many small businesses there are versus bigger multinational companies.

The question leading up to the sixth visualization asks, “How does payroll and revenue differ between black male and black female firm owners?” In order to create this visual a data frame where race is set to ‘black or African American’, ethnicity and veteran status is set to ‘total’, and firm size is set to ‘all firms’ is necessary. After this only male and female sex should be included. Finally, all variables except sex, payroll, and revenue should be filtered out. This final data frame is leveraged to create a double bar chart via the matplotlib library.



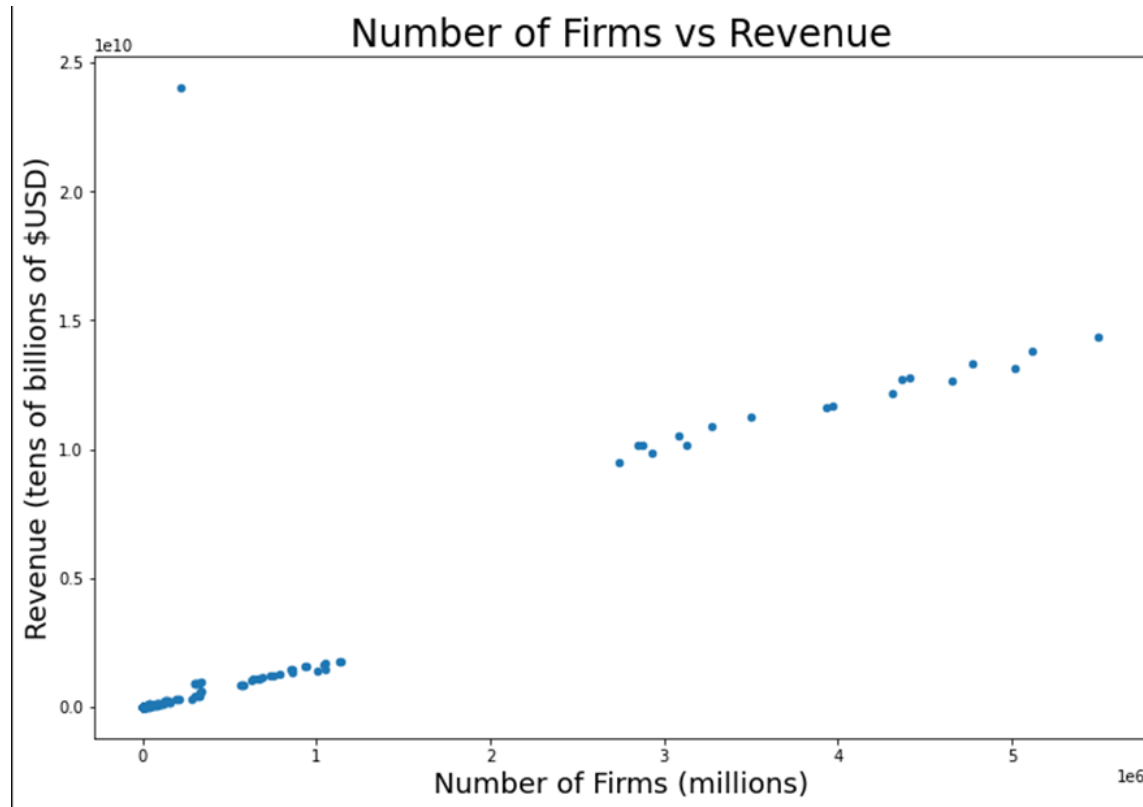
The graph above displays sex on the x axis and dollars in thousands on the y axis to quantify payroll and revenue. This visualization shows that firm owners who are black men tend to have a slightly higher payroll however earn a significantly higher amount of revenue compared to their female counterparts. These findings would be interesting to see if these proportions are consistent across different races and ethnicities.

The question leading up to the seventh visualization asks, “How does payroll and revenue differ between asian male and asian female firm owners?” In order to create this visual a data frame where race is set to ‘Asian’, ethnicity and veteran status is set to ‘total’, and firm size is set to ‘all firms’ is necessary. After this only male and female sex should be included. Finally, all variables except sex, payroll, and revenue should be filtered out. Using the matplotlib library a grouped bar chart was made.



The above graph displays sex on the x axis and dollars in thousands on the y axis to quantify payroll and revenue. The visualization gives a similar result to the previous one, with male asian business owners having just over double the payroll but significantly more revenue than female asians. This suggests the idea presented previously that this trend may be seen across all races and ethnicities is true.

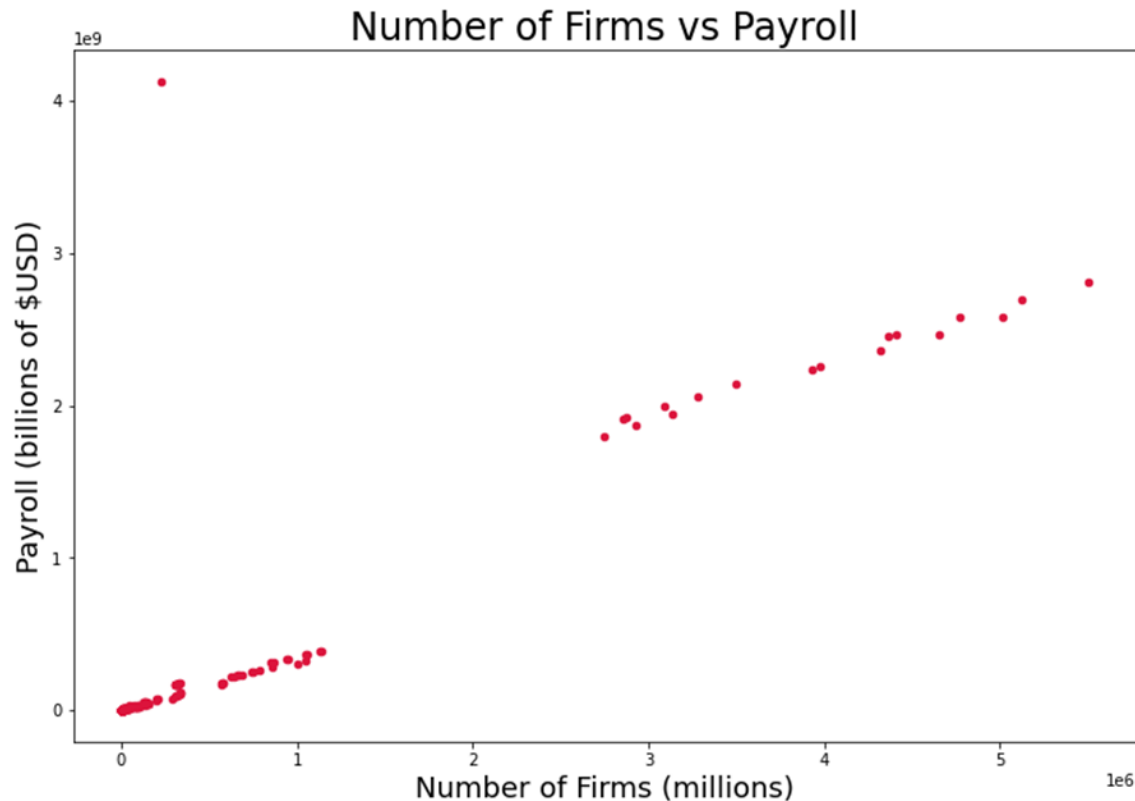
Moving forward, a goal was to analyze and compare how the number of firms correlates with the revenue and payroll using every combination of demographics produced. In the first graph shown below, the question created was “What is the correlation or relationship between the total number of firms per demographic combination and revenue of those firms?” In order to answer this question, the data frame needed to be narrowed down to get the records necessary. The decision was made to filter the data by “Firm Size” to only display the records that showed information containing the value “All firms”. Next, this was stored in a variable called “f” to be able to call this filtered data frame whenever needed. This process manufactured about 200 records that had different demographic combinations. Once those records were obtained, the next step was to create a scatter plot that compared the “Number of Firms” column to the “Revenue” column. The graph is shown below:



This graph shows the Number of Firms, listed in millions, on the x-axis and the corresponding Revenue of those firms, listed in tens of billions of \$USD, on the y-axis. Each data point on this scatter plot represents a different combination of the demographics within the data frame and its total revenue for those firms. As you can see the data displays a positive correlation between the two columns. To simplify it, this graph shows that as the number of firms increases, so will the amount of revenue for those firms.

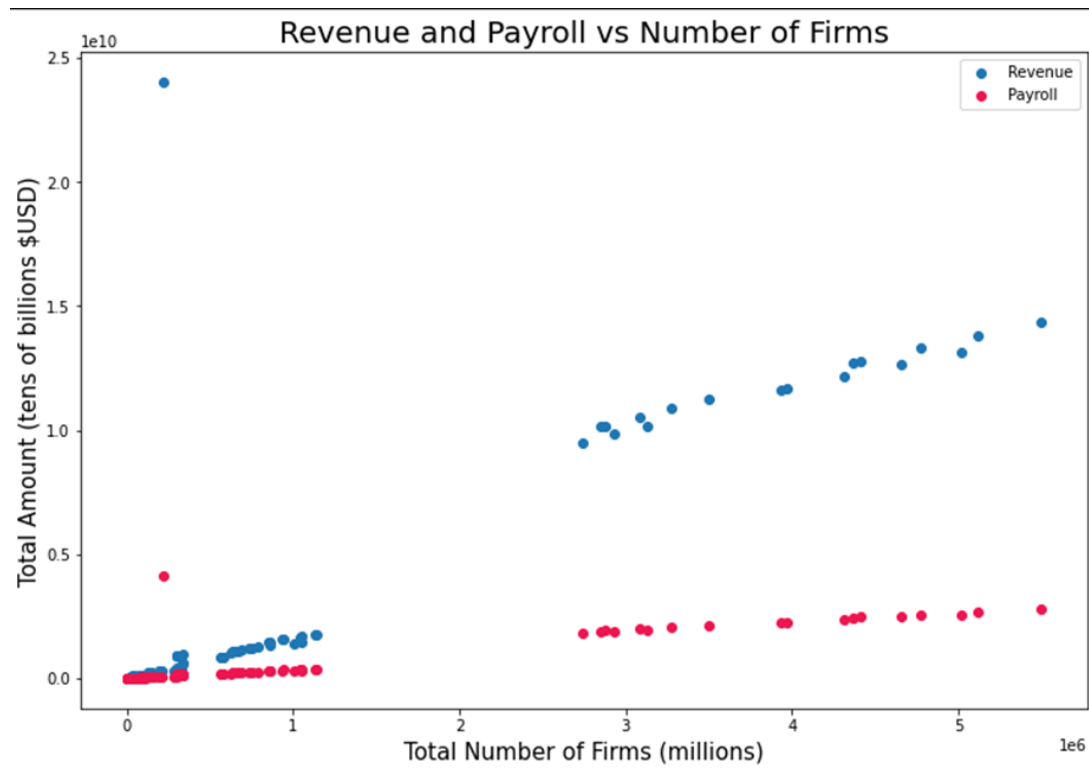
Similar to the previous graph, we were looking to answer the following question: "What is the correlation/relationship between the number of firms and the corresponding payroll of those of firms? Because the same data is being used, as we did for the prior scatter plot, there was no need to filter anything more than what was already done in

the filtering process. We simply switched columns from revenue to payroll, and created another scatter plot to display our findings. The graph is shown below:



This graph is displaying the total number of firms on the x-axis, and the corresponding payroll of those firms. Each datapoint on this graph is representing a different combination of demographics and its matching payroll. As expected, the graph displays a positive correlation/relationship between the two columns, like the previous scatter plot. In fact, these graphs are almost identical in a correlation aspect. They differ since payroll has much smaller values when compared to revenue. Revenue was displayed in tens of billions of \$USD, while payroll is displayed in billions of \$USD.

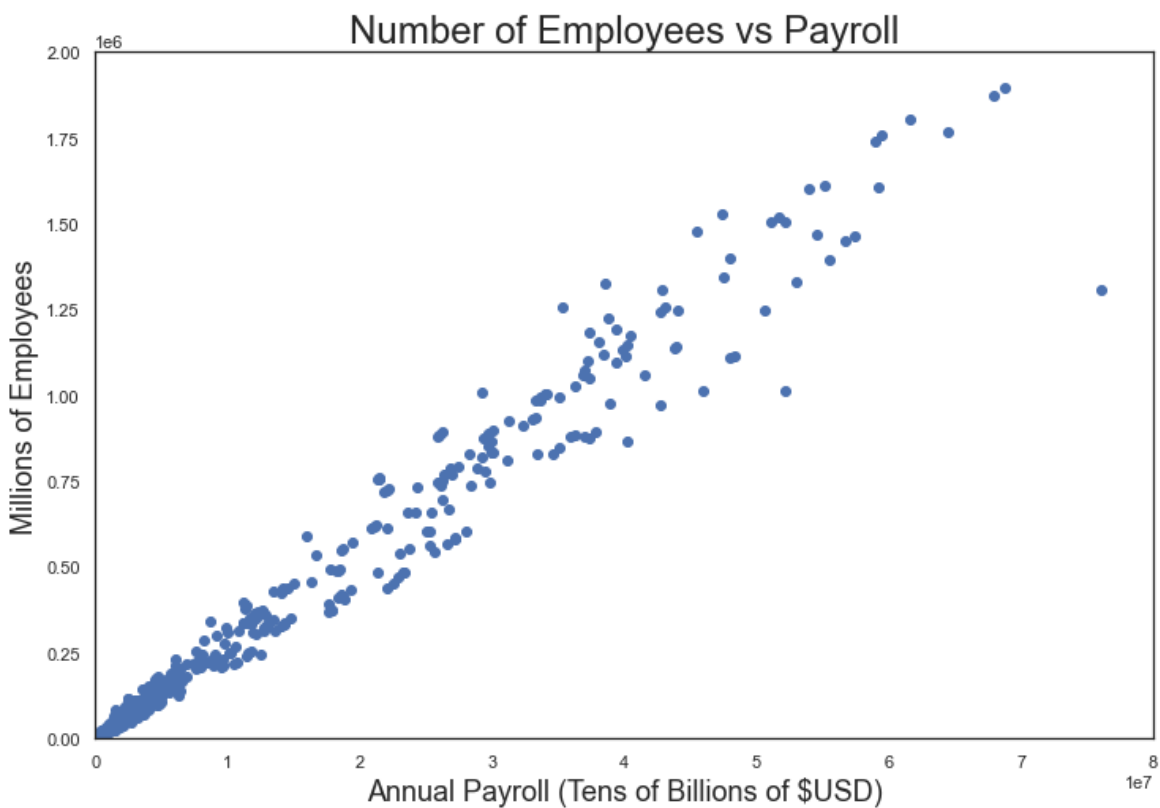
To further display the difference in the amounts, and to cure our own curiosity, we decided to create a multi-series with both revenue and payroll being compared to the number of firms. The graph is depicted below:



As you can see, the amount of revenue is excessively larger than that of payroll. The largest amount of payroll comes in at about 4.5 billion dollars; on the other hand, the largest amount of revenue comes in at just a shade under 25 billion dollars. Another detail this graph shows are the correlations. From this graph it is shown that "Revenue vs Number of firms" has a higher correlation than "Payroll vs Number of firms", as the trend of payroll is much more horizontal than that of revenue. This is quite misleading; prior to this graph being created, when solely payroll was plotted against a number of firms, the correlation was just as strong as that of revenue. What is skewing this graph and creating this misconception, is the excessive difference in the size of the amounts

between revenue and payroll. In summary, when being compared to the number of firms per demographic combination, both revenue and payroll have a positive correlation with the number of firms. Revenue is just increasing at a higher rate than payroll.

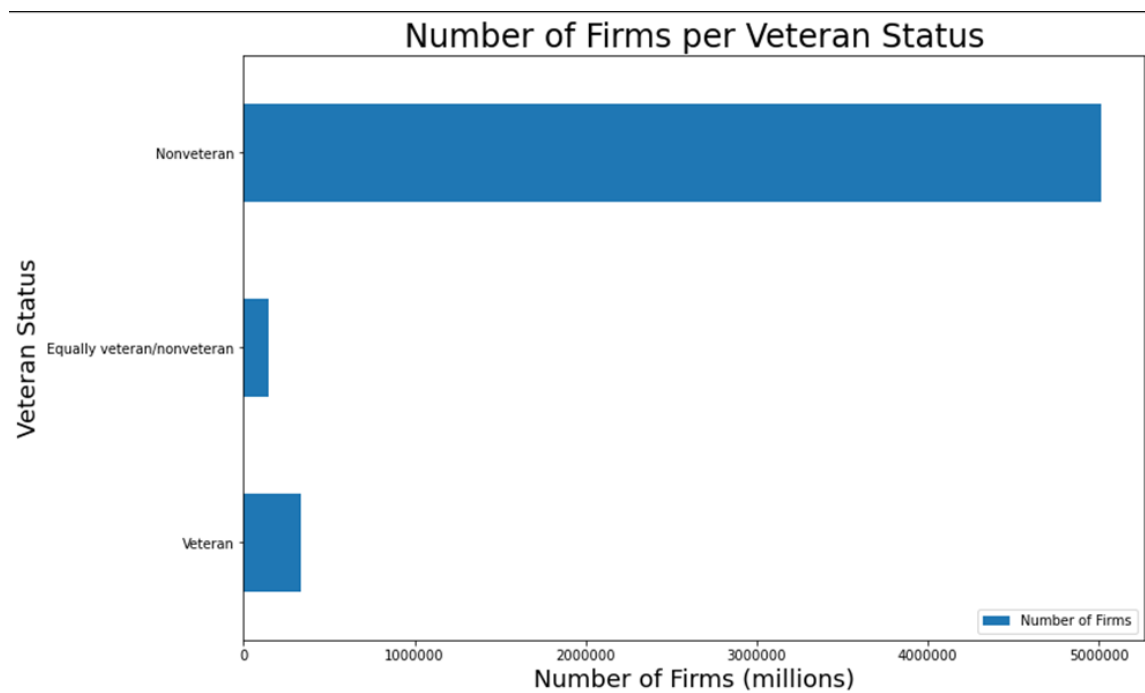
In addition to considering the number of firms as it correlates to payroll and revenue, another question was how the number of employees would affect the payroll. The following graph was made to visualize this question by using the matplotlib library. All firms and all races and ethnicities are included for this visualization, and overall totals are excluded to ignore outliers that would be present in the data otherwise.



The axes of this graph include the number of employees in millions of employees and annual payroll in tens of billions of USD. The hypothesis was that payroll and

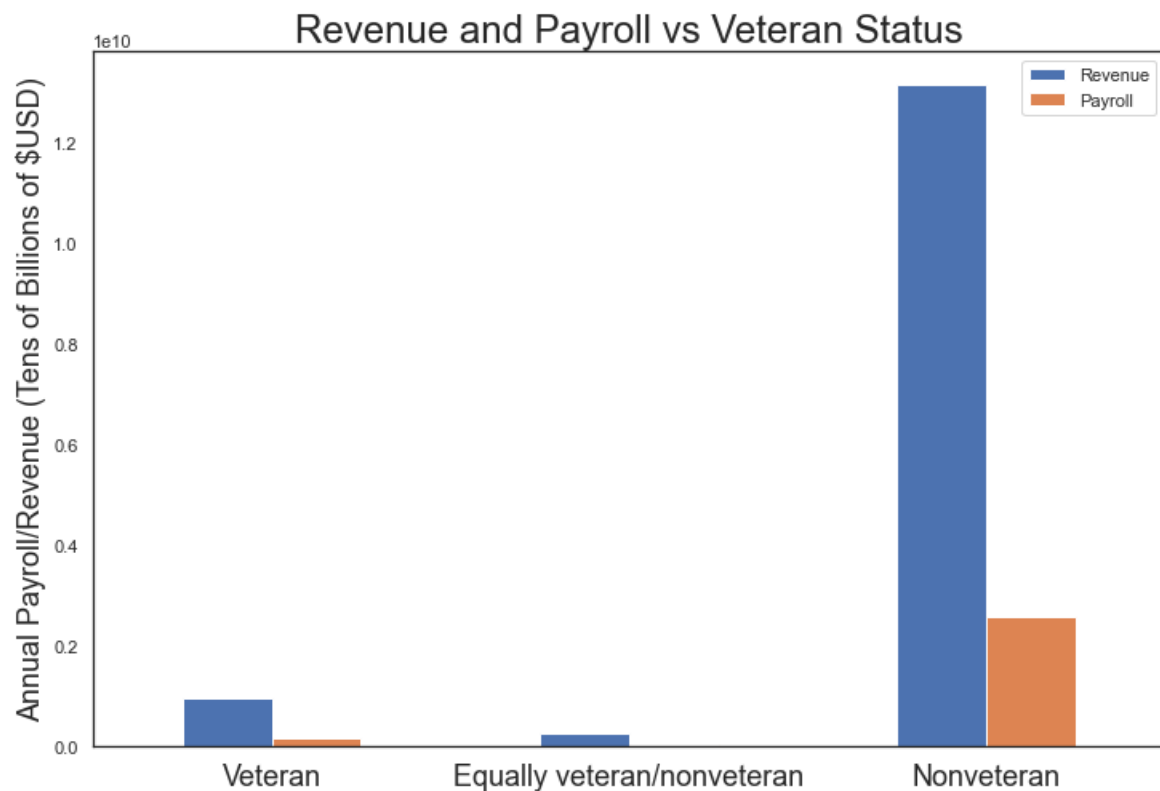
number of employees would be positively correlated since more employees would lead to more wages needing to be paid. This scatter plot confirmed this hypothesis by showing an obvious positive slope between the data.

Another question that was developed as we analyzed the data frame was, “How do the number of firms of different veteran statuses of all other demographics differ?” To answer this question, a variable needed to be made that stored the total Nonveteran, Equally veteran/nonveteran, and Veteran business owners with all other demographics such as Sex, Race, and Ethnicity set to their totals. In this variable the Firm Size was filtered to All firms in order to get the total Number of Firms. By doing this, the total number of firms of all three veteran statuses were displayed, making it quite easy to compare and analyze them. Using the veteran statuses and the “Number of Firms” column, a horizontal bar chart was created using the Matplotlib library to plot the data. The bar chart is depicted below:



The graph shows the three different veteran statuses, Nonveteran, Equally veteran/non-veteran, and Veteran on the y-axis, and the corresponding number of firms, by the millions, on the x-axis, ranging from 0 to 5,000,000 firms. The number of firms with an owner that has a Nonveteran status has the overwhelming majority of firms, coming in at around 5,000,000 firms (5M). Equally veteran/non-veteran business owners have the least number of firms coming in at around 145,000. Veteran business owners come in at around 340,000. This graph shows us that the clear majority of business owners have a veteran status of Nonveteran.

The next visualization addresses the question “How do payrolls and revenues vary with each veteran status?” Using the same filtering as the previous visualization, a graph was made to answer this question using the matplotlib library.



The conclusions that can be drawn from the above graph are that the payroll is much higher for nonveterans compared to veterans or businesses that are equally split and the revenue follows a similar trend. These observations are expected based on the previous graph which shows that there are many more firms owned by nonveterans, resulting in higher revenue and payroll.

Given the findings from this analysis there are plenty of conclusions which can be drawn about how a firm owner's demographics may or may not influence the firm's financials. While some visualizations offered a concise answer to a specific question others sparked questions to promote deeper analysis into a specific topic. Furthermore, there were hundreds of demographic combinations that had an effect on multiple aspects of their corresponding firms. The main factors being Sex, Ethnicity, Race, and Veteran status, one can see that based on the status of each demographic, there can be great differences in the monetary aspects of the firms involved. After analyzing our data and visualizations, here are the answers we were able to find: Male business owners have much higher annual payrolls and revenues than female business owners regardless of race, White business owners dominate in all business financial data, the bigger the firm size the higher the payroll and revenue, there are more small firms than large firms, there is a strong positive correlation between number of employees and payroll, more non-veterans own businesses than veterans, and non-veteran business owners have a higher payroll and revenue than veteran business owners. All in all there is no denying that an owner's demographics has some form of influence on the firm.